NCER National Centre for Econometric Research

NCER Working Paper Series

Testing for Nonlinearity in Mean in the Presence of Heteroskedasticity

R. Becker and S. Hurn

Working Paper #8

January 2007

Testing for nonlinearity in mean in the presence of heteroskedasticity

R. Becker* and A. S. Hurn#

*Economics, School of Social Sciences, University of Manchester,

#School of Economics and Finance, Queensland University of Technology

Abstract

This paper considers an important practical problem in testing time-series data for nonlinearity in mean. Most popular tests reject the null hypothesis of linearity too frequently if the the data are heteroskedastic. Two approaches to redressing this size distortion are considered, both of which have been proposed previously in the literature although not in relation to this particular problem. These are the heteroskedasticity-robust-auxiliary-regression approach and the wild bootstrap. Simulation results indicate that both approaches are effective in reducing the size distortion and that the wild bootstrap offers better performance in smaller samples. Two practical examples are then used to illustrate the procedures and demonstrate the potential pitfalls encountered when using non-robust tests.

Keywords

nonlinearity in mean, heteroskedasticity, wild bootstrap, empirical size and power

JEL Classification C12, C22, C52

Corresponding author

Ralf Becker <ralf.becker@manchester.ac.uk>

1 Introduction

The empirical properties of econometric tests for nonlinearity in mean in time-series data have been well documented in the literature (Lee *et al.*, 1993, Teräsvirta *et al.,* 1993 and Barnett et al.*,* 1997). As a result it is reasonably well known that non-constant variance in the data can cause problems for these tests. In particular, there is a tendency to over-reject the null hypothesis of linearity in mean if the time series being tested is heteroskedastic. In some cases, however, it may be important to establish whether the rejection of the null hypothesis is due to neglected nonlinearity in mean or whether the rejection is merely a consequence of heteroskedastic data. An example of one such instance is the growing literature on the importance of asymmetric loss functions in the context of the conduct of monetary policy (Kim *et al.*, 2002, Elliot *et al.*, 2003). This type of loss function could imply a nonlinear policy reaction function but testing for nonlinearity will be complicated by the fact that the macroeconomic time series data used in the estimation is likely to be heteroskedastic. To ascertain whether the policy reaction rule should correctly be specified as nonlinear function, a test for nonlinearity in mean is required that has the correct size even when heteroskedasticity is present.

The basic test for nonlinearity that will be used in this paper is a version of the neural network test (White, 1989) as implemented by Teräsvirta *et al.* (1993) and known as the V23 test[1]. The strategy, which is proposed here, is to use the heteroskedastic-robust-regression framework outlined by Davidson and MacKinnon (1985) to correct for the size distortion suffered by tests for nonlinearity in mean in the presence of heteroskedasticity . Clear expositions of this method in the context of testing for nonlinearity can also be found in Granger and Teräsvirta (1993). Although this framework is capable of dealing with the potential presence of heteroskedasticity of an unknown form, the form of non-constant variance used in the simulation design is limited to the autoregressive conditional heteroskedasticity, or (G)ARCH, model introduced by Engle (1982) and Bollerslev (1986). This choice is deter-

[1] This test is used because of its suitability for use in the proposed testing strategy and also because it is known to have good power against a number of nonlinear models. Of course the testing framework outlined in the paper can be used in conjunction with other well-known tests for nonlinearity.

mined by the prevalence of (G)ARCH in economic and financial time-series data. In addition to outlining the heteroskedastic-robust testing strategy in the context of testing for nonlinearity in mean, an appropriate bootstrapping technique will be investigated to see if this improves the small sample performance of the test.

This paper makes a number of contributions to the current state of the literature, both theoretical and empirical. In terms of theory, the paper establishes the conditions that need to be imposed on a stochastic process to ensure that the heteroskedastic-robust version of the V23 test has the correct asymptotic distribution. In addition, it is shown that the wild bootstrap proposed by Liu (1988), Mammen (1993) and Davidson and Flachaire (2000) can be used to provide a consistent estimate of this asymptotic distribution. Important groundwork with respect to the applicability of bootstrapping techniques to (G)ARCH processes has been provided by Gonçalves and Kilian (2004) who establish process conditions required to achieve consistent inference on AR parameters. These results will be used and extended to suit the robust-regression testing framework of this paper. From an empirical perspective, the size and power of the robust-regression approach to testing for nonlinearity in mean in the presence of non-constant variance are evaluated. This testing framework is shown to be an effective way of reducing the size distortion suffered by conventional testing procedures. It is also demonstrated that the use of the wild bootstrap, in conjunction with the robust-regression approach offers further small-sample improvements to the size of the test. This result derives from the ability of the wild bootstrap to replicate the lower-order moments of the empirical distribution of residuals.

The rest of the paper is structured as follows. Section 2 is a brief perspective on the testing problem that covers the heteroskedasticity-robust auxiliary regression approach to testing for nonlinearity in mean and also introduces the wild bootstrap. Section 3 establishes the required theoretical results. In Sections 4 and 5 of the paper the empirical performance of the auxiliary regression using both the asymptotic distribution and the wild bootstrap to determine the significance of the testing procedure is evaluated. In Section 6 the tests are applied to the Yen/US$ exchange rate and the US 3-month Treasury Bill rate, being two of the data sets examined by Lee *et al.* (1993) in their comprehensive comparison of tests for nonlinearity. It is shown that ignoring the presence of heteroskedasticity can result in the null of linearity in mean being rejected too easily. Section 7 is a brief conclusion.

2 An overview of the testing problem

2.1 The V23 test

Consider the nonlinear time-series model

$$y_t = g\left(\mathbf{y}_{t-1};\boldsymbol{\beta}\right) + \varepsilon_t \qquad \varepsilon_t \sim i.i.d.\left(0, h^2\right) \tag{1}$$

where the mean function $g\left(\mathbf{y}_{t-1};\boldsymbol{\beta}\right)$ may be decomposed into linear and nonlinear components as follows

$$y_t = \alpha + \mathbf{y}_{t-1}\boldsymbol{\beta} + \phi\left(\mathbf{y}_{t-1};\boldsymbol{\delta}\right) + \varepsilon_t. \tag{2}$$

Note that $\mathbf{y}_{t-1} = \left(y_{t-1}, \ldots, y_{t-p}\right)'$ which implies that the analysis is restricted to time-series models[2]. In this specification α is a scalar and $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are $(p \times 1)$ and $(q \times 1)$ parameter vectors representing the linear and nonlinear contributions to the mean respectively. The definition of linearity in mean is $\phi\left(\mathbf{y}_{t-1};\boldsymbol{\delta}\right) = 0$. For most nonlinear models $\phi\left(\mathbf{y}_{t-1};\boldsymbol{\delta}\right)$ can be reformulated so that $\phi\left(\mathbf{y}_{t-1};\boldsymbol{\delta}\right) = 0$ if $\boldsymbol{\delta} = 0$. Sometimes it is even sufficient if one particular parameter in the vector $\boldsymbol{\delta}$ equals zero although this situation introduces the problem of the remaining parameters in $\boldsymbol{\delta}$ being unidentified under the null hypothesis.

In order to implement a test for nonlinearity in this framework, the form of the function $\phi\left(\mathbf{y}_{t-1};\boldsymbol{\delta}\right)$ must be specified, reflecting the nonlinear model that is envisaged under the alternative hypothesis. A popular specification for $\phi\left(\mathbf{y}_{t-1};\boldsymbol{\delta}\right)$ that has good power against a range of alternative nonlinear models uses second- and third-order cross products of elements in $\mathbf{y}_{t-1}$. This specification is a variant, introduced by Teräsvirta *et al.* (1993), of the original test proposed by White (1989) where the nonlinear model under the alternative hypothesis takes the form of a neural network. The test regression is

$$y_t = \mathbf{y}_{t-1}\boldsymbol{\beta} + \sum_{i=0}^{p}\sum_{j=i}^{p}\delta_{ij}y_{t-i}y_{t-j} + \sum_{i=1}^{p}\sum_{j=i}^{p}\sum_{k=j}^{p}\delta_{ijk}y_{t-i}y_{t-j}y_{t-k} + \varepsilon_t, \tag{3}$$

[2] Extending the analysis to models including exogneous variables is straightforward.

and the associated null hypothesis of linearity is specified as

$$H_0 : \delta_{ij} = \delta_{ijk} = 0 \qquad \forall\ i, j, k$$

The test can also be conducted within a Lagrange Multiplier framework as follows. Let $\widehat{\boldsymbol{\beta}}$ be a consistent estimate of the parameter vector under the null hypothesis of linearity and let the scores with respect to the parameter vector $\boldsymbol{\delta}$ be denoted

$$\hat{\mathbf{S}}\left(\hat{\boldsymbol{\beta}}\right) = \frac{1}{T}\sum \frac{\partial L_t}{\partial \boldsymbol{\delta}}.$$

If the relation is indeed linear then $\hat{\mathbf{S}}\left(\hat{\boldsymbol{\beta}}\right)$ should be close to zero, and the LM test of this hypothesis is given by

$$LM = T\ \hat{\mathbf{S}}'\mathbf{I}\left(\hat{\mathbf{S}}\right)^{-1}\hat{\mathbf{S}}. \tag{4}$$

where the covariance of the scores is the information matrix, $\mathbf{I}\left(\hat{\mathbf{S}}\right)$.

It is well known that this LM statistic is easily computed by means of an auxiliary regression (see for example, MacKinnon, 1992, p109). Define $\widehat{\varepsilon}_t$ as the residuals estimated from the linear model

$$y_t = \mathbf{y}_{t-1}\boldsymbol{\beta} + \varepsilon_t. \tag{5}$$

The LM test statistic may be computed as

$$LM = TR^2. \tag{6}$$

where the coefficient of determination R^2 is calculated from the auxiliary regression which regresses the residual estimates $\widehat{\varepsilon}_t$ on the explanatory variables and the partial derivatives of $\phi_t = \phi\left(\mathbf{y}_{t-1}; \boldsymbol{\delta}\right)$ with respect to the parameter vector $\boldsymbol{\delta}$. The regression is given by

$$\begin{aligned} \widehat{\varepsilon}_t &= \mathbf{z}_t\boldsymbol{\theta} + \nu_t \\ \mathbf{z}_t &= \left(\mathbf{y}'_{t-1}, \mathbf{d}'_t\right)' \end{aligned} \tag{7}$$

where the vector $\mathbf{d}_t$ is defined as

$$\mathbf{d}_t = \partial\phi_t / \partial\boldsymbol{\delta}. \tag{8}$$

Essentially the vector $\mathbf{d}_t$ comprises all the unique second- and third-order cross products of $\mathbf{y}_{t-1}$.

2.2 Heteroskedastic-robust regression

To this point the innovations, ε_t, of the nonlinear time-series model of equation (1) have been assumed to be *i.i.d.* This assumption is now relaxed to allow for heteroskedasticity as follows

$$y_t = g(\mathbf{y}_{t-1}; \boldsymbol{\beta}) + h(\mathbf{y}_{t-1}; \boldsymbol{\gamma})\,\varepsilon_t \qquad \varepsilon_t \sim i.i.d.\,(0,1)\,. \tag{9}$$

The extension to non-constant residual variance $h^2(\mathbf{y}_{t-1}; \boldsymbol{\gamma})$ is straightforward, so long as $\partial h^2(\mathbf{y}_{t-1}; \boldsymbol{\gamma}) / \partial \boldsymbol{\beta} = \mathbf{0}$, and $h^2(\mathbf{y}_{t-1}; \boldsymbol{\gamma})$ is completely specified. The LM test for nonlinearity may now be implemented as follows. Estimate the residuals from the heteroskedastic model imposing the null hypothesis of linearity. This will provide estimates of the standardized residuals $\widetilde{\varepsilon}_t = \widehat{\varepsilon}_t \hat{h}^{-1}$ and also $\widetilde{\mathbf{y}}_{t-1} = \mathbf{y}_{t-1}\hat{h}^{-1}$ and $\tilde{\mathbf{d}}_t = \mathbf{d}_t \hat{h}^{-1}$. The LM test is again calculated as TR^2, with R^2 now being the coefficient of determination from the auxiliary regression with the standardised variables (Granger and Teräsvirta, 1993).

Often, however, not enough information is available to specify the variance function $h^2(\mathbf{w}_t; \boldsymbol{\gamma})$ and it is desirable to cater for unspecified heteroskedasticity when testing the linearity of the conditional mean. A regression-based approach to achieve this was proposed by Davidson and MacKinnon (1985). It is best described by the following steps

1. Estimate the restricted residuals $\widehat{\varepsilon}_t$ from equation (5).

2. Let $\mathbf{D}$ be the $(T \times q)$ matrix of the stacked vectors $\mathbf{d}_t$. Regress the q elements in $\mathbf{D}$ individually on $\mathbf{Y}$, which is the $(T \times p)$ matrix of the stacked vectors $\mathbf{y}_{t-1}$. Save the q resulting residual vectors $\mathbf{r}(\mathbf{D}^j\,|\mathbf{Y})$, where $\mathbf{D}^j$ indicates the jth column $(j = 1, \ldots, q)$ of $\mathbf{D}$.

3. Compute the weighted residuals $\tilde{r}_t(\mathbf{D}^j\,|\mathbf{Y}) = r_t(\mathbf{D}^j\,|\mathbf{Y})\,\widehat{\varepsilon}_t$ and

4. Regress a $(T \times 1)$ vector of ones, $\mathbf{1}$, on the q regressors computed in step 3, $\tilde{\mathbf{r}}(\mathbf{D}^j\,|\mathbf{Y})$, $j = 1, \ldots, q$. The test statistic is computed as the explained sum of squares or $T - RSS$ from this regression[3].

[3] This is numerically equivalent to calculating the test statistic $\widehat{\boldsymbol{\varepsilon}}'\mathbf{M}_Y\mathbf{D}\left(\mathbf{D}'\mathbf{M}_Y\widehat{\Omega}\mathbf{M}_Y\mathbf{D}\right)^{-1}\mathbf{D}'\mathbf{M}_Y\widehat{\boldsymbol{\varepsilon}}$ as in

Wooldridge (1990) provides the conditions under which this testing procedure will result in a test statistic whose asymptotic distribution under the null hypothesis is χ^2_q . As previously noted, in the context of the V23 test, the vector $\mathbf{d}_t$ comprises all the unique second- and third-order cross products of the lagged dependent variable. It follows, therefore, that this test will have the required asymptotic distribution if and only if the process $\{y_t\}$ satisfy the conditions established by Wooldridge (1990). The existence of the asymptotic distribution for the V23 test is established in Section 3.

2.3 Wild bootstrap

Recent work by Godfrey and Orme (2004) has indicated a persistent small-sample size distortion in the heteroskedastic-robust testing framework. The heteroskedastic processes of interest here, namely (G)ARCH processes, are known to be near epoch dependent (NED) functions of mixing processes (Sin and White, 1996, Davidson, 2000). It has recently been established that the block (Künsch, 1989), stationary bootstrap (Politis and Romano, 1994) and the pair bootstrap (Gonçalves and Kilian, 2004) deliver consistent inference on parameter estimates when applied in the context of NED processes (Gonçalves and White, 2000, 2001). Unfortunately this result is not useful in the present context, since these bootstrap techniques not only preserve the structure in the residual variance but will potentially also capture nonlinear dependence in the data[4]. They are, therefore, not suitable in the context of testing for nonlinear dependence in mean where the null hypothesis is that of linearity, as it is paramount that the bootstrapping technique complies with the restrictions imposed by the null hypothesis. In these circumstances the wild bootstrap proposed by Liu (1988), Mammen (1993) and Davidson and Flachaire (2000) appears to be the only suitable alternative. The wild bootstrap has been shown to be particularly useful in bootstrapping (G)ARCH processes (Gonçalves and Kilian, 2004). In conjunction with the robust-regression approach, therefore, the wild bootstrap may offer improvements to the size of the V23 test in small

Davidson and MacKinnon (1985) and Godfrey and Orme (2004). The presentation here is in terms of the auxilliary regression approach which can be implemented easily in standard econometric packages.

[4] All these three bootstrap techniques resample pairs of $(y_t, \mathbf{y}_{t-1})$ from the original samples. Any relationship between y_t and $\mathbf{y}_{t-1}$, linear and nonlinear, is conserved by such a resampling scheme.

samples, given that the heteroskedasticity-robust V23 test statistic is asymptotically pivotal.

The intuition of the wild bootstrap is to preserve the observed time pattern in the residual variance. This is achieved by resampling the residuals in such a way that (at least) the first two moments of the observed regression residuals are maintained. Consider the residuals $\{\widetilde{\varepsilon}_t\}, t = 1 \ldots T$ defined by $\tilde{\varepsilon}_t = \delta\hat{\varepsilon}_t - \bar{\varepsilon}$, where $\hat{\varepsilon}_t$ is the OLS residual from the model estimated under the null hypothesis and $\bar{\varepsilon} = 1/T \sum \delta\hat{\varepsilon}_t$. The rescaling factor is $\delta = \sqrt{T/(T-k)}$ with k being the number of estimated parameters (Mammen, 1993, Flachaire, 1999, Bergström, 1999). The general resampling scheme for the wild bootstrap is given by

$$\varepsilon_t^* = v_t \cdot g(\tilde{\varepsilon}_t) \tag{10}$$

where $g(\tilde{\varepsilon}_t) = |\tilde{\varepsilon}_t|$ and

$$v_t = \begin{cases} 1 & \textit{with probability } 0.5 \\ -1 & \textit{with probability } 0.5 \end{cases}. \tag{11}$$

which is the algorithm suggested by Davidson and Flachaire (2000). The limited simulation evidence provided by Davidson and Flachaire (2000) and Godfrey and Orme (2001, 2004) tends to support these choices for $g(\tilde{\varepsilon}_t)$ and v_t. Indeed, these choices are preferred to an alternative resampling scheme (Mammen, 1993), which also preserves the third moment of the observed regression residuals.

There are two methods for creating bootstrap samples using the wild bootstrap in the context of autoregressive models, namely, the *fixed-design wild bootstrap*, $FDWB$, and the *recursive wild bootstrap*, RWB (Gonçalves and Kilian, 2004). The former generates the bootstrap realizations from

$$y_t^* = \mathbf{y}_{t-1}\hat{\boldsymbol{\beta}} + \varepsilon_t^*,$$

whereas the latter requires the recursive scheme

$$y_t^* = \mathbf{y}_{t-1}^*\hat{\boldsymbol{\beta}} + \varepsilon_t^*.$$

For the latter starting values for $\mathbf{y}_0^*$ are required. In order to negate any significant impact of the choice of starting values a series that is longer than required is generated and then the initial redundant observations discarded.

Gonçalves and Kilian (2004) show that both versions of the wild bootstrap (recursive and fixed-design) allow consistent inference over the regression parameter vector $\boldsymbol{\beta}$, when the data follow a range of (G)ARCH processes and both methods will be used in the simulation evidence presented Sections 4 and 5 of this paper.

3 Theoretical results

This section will establish two important theoretical results. *First*, the set of conditions on the data generating process $\{y_t\}$ will be established that ensures that the heteroskedastic-robust version of the V23 test has an asymptotic χ_q^2 distribution. *Second*, it will be shown that the wild bootstrap will generate a consistent estimate of this distribution.

3.1 Asymptotic distribution of the V23 test

When imposing the null hypothesis, the data generating process for $\{y_t\}$ is

$$y_t = \mathbf{y}_{t-1}\boldsymbol{\beta}_l + \varepsilon_t \tag{12}$$

where, as before, $\mathbf{y}_{t-1} = (y_{t-1}, \ldots, y_{t-p})'$. An alternative formulation is

$$\beta(L)\, y_t = \varepsilon_t$$

where the lag polynomial $\beta(L)$ is assumed to have all roots outside the unit circle. In order to prove the theoretical results required for consistent inference by the V23 test, a number of assumptions on the residual sequence $\{\varepsilon_t\}$ are required and these are stated in Assumptions A (A1 - A6).

Assumption (A1) $E(\varepsilon_t \,|\mathcal{F}_{t-1}) = 0$, almost surely, where $\mathcal{F}_{t-1} = \sigma(\varepsilon_{t-1}, \varepsilon_{t-2}, ...)$ is the σ-field generated by $\{\varepsilon_{t-1}, \varepsilon_{t-2,...}\}$.

Assumption (A2) $E(\varepsilon_t^2) = \sigma^2 < \infty$

Assumption (A3) $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(\varepsilon_t^2 \,|\mathcal{F}_{t-1}) = \sigma^2 > 0$ in probability.

Assumption (A4) $E\left(\varepsilon_{t-k_1}\cdot\ldots\cdot\varepsilon_{t-k_i}\right)=\sigma^i\tau_{k_1,\ldots,k_i}$ for any $0\leq k_1\leq k_2\leq\ldots\leq k_{i-1}\leq k_i$ and all t is uniformly bounded for $i=3,\ldots,8$. When $k_1<k_2$ then $\tau_{k_1,\ldots,k_i}=0$ due to A1.

Assumption (A5) $\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}E\left[\varepsilon_{t-k_1}\cdot\ldots\cdot\varepsilon_{t-k_i}|\mathcal{F}_{t-k_1-1}\right]=\sigma^i\tau_{k_1,\ldots,k_i}$ in probability for any $0\leq k_1\leq k_2\leq\ldots\leq k_{i-1}\leq k_i$, for $i=3,\ldots,8$.

Assumption (A6) $E\left|\varepsilon_t\right|^{16r}$ is uniformly bounded, for some $r>1$.

In essence these conditions require that the higher order moments of the residual sequence are well behaved and it has been demonstrated by Deo (2000) that a number of (G)ARCH and stochastic volatility models satisfy these assumptions, conditional on innovations which possess an appropriate number of higher moments.

The relevant asymptotic theory for the V23 test is adapted from the relevant theorem by Wooldridge (1990). Let $\boldsymbol{\lambda}_t\left(\mathbf{y}_{t-1}\right)=\mathbf{d}_t$ represent a q-dimensional vector of unique second and third-order cross products of the p elements in $\mathbf{y}_{t-1}$, which does not depend on any parameter estimates or other nuisance parameters, thus simplifying the analysis considerably. Furthermore let

$$\boldsymbol{\mu}_t\left(\mathbf{y}_{t-1},\widehat{\boldsymbol{\beta}}\right)\equiv E\left[\left(\frac{\partial\left[\varepsilon_t\left(y_t,\mathbf{y}_{t-1},\widehat{\boldsymbol{\beta}}\right)\right]}{\partial\boldsymbol{\beta}}\right)|\mathbf{y}_{t-1}\right]=-\mathbf{y}_{t-1}$$

where the last equality follows from the linearity of the model under the null hypothesis. Wooldridge's Theorem 2.1 can now be restated as follows.

Theorem 1 (Wooldridge 2.1) *Assume the following conditions hold under the null hypothesis:*

(i) Regularity conditions A.1 (Wooldridge, 1990, Mathematical Appendix, p 40)

(ii) For some $\boldsymbol{\beta}_0\in int\left(\Phi\right),$

(a) $E\left[\varepsilon_t\left(y_t,\mathbf{y}_{t-1},\boldsymbol{\beta}_0\right)|\mathbf{y}_{t-1}\right]=0,\quad t=1,2,\ldots,T;$

(b) $T^{1/2}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)=O_p\left(1\right).$

Then

$$\widetilde{\xi}_T = T^{-1/2} \sum_{t=1}^{T} \left[\boldsymbol{\lambda}_t^0 - \boldsymbol{\mu}_t^0 \mathbf{B}_T^0 \right] \varepsilon_t^0 + o_p(1)$$

where

$$B_T^0 \equiv \left(\sum_{t=1}^{T} E\left[\boldsymbol{\mu}_t^{0\prime} \boldsymbol{\mu}_t^0 \right] \right)^{-1} \sum_{t=1}^{T} E\left[\boldsymbol{\mu}_t^{0\prime} \boldsymbol{\lambda}_t^0 \right].$$

Further,

$$T R_u^2 \xrightarrow{d} \chi_q^2,$$

where R_u^2 is the uncentered R^2 from the regression

$$\mathbf{1} \text{ on } \widehat{\varepsilon}_t' \left[\widehat{\boldsymbol{\lambda}}_t - \widehat{\boldsymbol{\mu}}_t \widehat{\mathbf{B}}_T \right]$$

estimating $\widetilde{\xi}_T$ and $\widehat{\mathbf{B}}_T$ from

$$\widetilde{\xi}_T = T^{-1/2} \sum_{t=1}^{T} \left[\widehat{\boldsymbol{\lambda}}_t - \widehat{\boldsymbol{\mu}}_t \widehat{\mathbf{B}}_T \right] \widehat{\varepsilon}_t$$

and

$$\widehat{\mathbf{B}}_T \equiv \left(\sum_{t=1}^{T} \widehat{\boldsymbol{\mu}}_t' \widehat{\boldsymbol{\mu}}_t \right)^{-1} \sum_{t=1}^{T} E\left[\widehat{\boldsymbol{\mu}}_t' \widehat{\boldsymbol{\lambda}}_t \right].$$

respectively.

Proof. See Wooldridge (1990, p 41). ■

Closer inspection reveals that the auxiliary regression outlined in this theorem is in fact identical to that outlined earlier in the context of the heteroskedastic-robust implementation of the V23 test, recognizing that $\left[\widehat{\boldsymbol{\lambda}}_t - \widehat{\boldsymbol{\mu}}_t \widehat{\mathbf{B}}_T \right]$ is the residual obtained by regressing the elements in $\widehat{\boldsymbol{\lambda}}_t$ on $\widehat{\boldsymbol{\mu}}_t = \mathbf{y}_{t-1}$. This theorem therefore provides the necessary proof of the existence of the asymptotic χ_q^2 distribution for the V23 test, provided that the conditions required by the theorem are satisfied in the current context. This is now established in the following Lemma.

Lemma 2 *Given assumptions A1 - A6, the conditions A.1 in Wooldridge (1990, p 40) are fulfilled.*

Proof. Appendix ■

3.2 Consistency of the wild bootstrap

Having established that the V23 test, in its robust implementation, has an asymptotic χ_q^2 distribution, it is now necessary to prove that wild bootstrap will provide a consistent estimate of this distribution. Gonçalves and Kilian (2004) have recently shown that both versions of the wild bootstrap (recursive and fixed-design) allow consistent inference over the regression parameter vector $\boldsymbol{\beta}$, when the data follow a range of (G)ARCH processes. The task here is to extend these results to approximating the asymptotic χ_q^2 distribution of the V23 test. The proof provided here applies only to the fixed-design wild bootstrap because stronger conditions are required for the recursive bootstrap and, in any event, the existing evidence suggests that there are no significant differences in the empirical performance of the two resampling schemes.

Theorem 3 *Under the assumptions A1 - A5 it follows that*

$$\sup_{x \in R^+} \left| P^* \left(TR_u^{2*} \leq x \right) - P \left(TR_u^2 \leq x \right) \right| \xrightarrow{p} 0$$

where P^* is the probability measure induced by the fixed design wild bootstrap. TR_u^2 and TR_u^{2*} are the robust regression test statistics based on the data and the fixed design wild bootstrap replications respectively.

The proof, provided in the Appendix, draws on Wooldridge's (1990) demonstration that

$$T^{-1/2} \sum_{t=1}^{T} \left(\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t \mathbf{B}_T \right)' \widehat{\varepsilon}_t \xrightarrow{d} N\left(0, \Xi_T\right).$$

In the context of the bootstrap what is required is the similar result that

$$T^{-1/2} \sum_{t=1}^{T} \left(\boldsymbol{\lambda}_t^* - \boldsymbol{\mu}_t^* \mathbf{B}_T^* \right)' \varepsilon_t^* \xrightarrow{d} N\left(0, \Xi_T\right)$$

where $\boldsymbol{\lambda}_t^*$, $\boldsymbol{\mu}_t^*$ and $\mathbf{B}_T^*$ refer to bootstrapped quantities. Of course, an important part of the proof is to establish that

$$\Xi_T^* \xrightarrow{p} \Xi_T^0.$$

Proof. Appendix ■

4 Design of the simulation experiments

The empirical size and power of the proposed testing strategy need to be assessed by means of a series of simple Monte Carlo experiments. Specifically, the data generating processes required to conduct the experiments fall naturally into two categories.

1. *Linear-in-mean processes (size simulations)*:

   The linear models used in the size simulations are: a time-series of standard normal random numbers (RNDN) and an autoregressive model of order one (AR1) are models which are both linear in mean and in variance; two ARCH and three GARCH models are models which are linear in mean but nonlinear in variance.

2. *Nonlinear processes (power simulations)*:

   The nonlinear models used in the power simulations are: a bilinear model (BILIN); a threshold autoregressive model (TAR); a sign autoregressive model (SAR); a nonlinear autoregressive model (NAR); a bilinear autoregressive model (BILINAR); and a logistic smooth transition autoregressive model (LSTAR).

In the simulations the null hypothesis is represented either by an autoregressive model of order one (RNDN, AR1, TAR, SAR, NAR, ARCH and GARCH) or of order two (BILIN, BILINAR and LSTAR). All linear models are estimated with a constant. The exact specifications are given in Appendix A. The sample size is set to be either 50, 100 or 200 and the size and power results are based on $5,000$ simulations. The bootstrap tests are applied with 400 bootstrap replications.

Size and *power* simulations were conducted using disturbances which are conditionally normally distributed. In order to investigate the sensitivity of the robust regression approach and the wild bootstrap approach, the *size* simulations were repeated using conditional standardized $\chi^2(2)$ and $t(5)$ random deviates. The results will supplement the empirical investigation by Godfrey and Orme (2001) who investigate the robustness of several wild bootstrap mechanisms to nonnormality. Both variations of the wild bootstrap, $FDWB$ and RWB, are applied.

Before proceeding to discuss the results of these simulation exercises it is necessary to examine whether or not the stringent moment requirements for the asymptotic theory as outlined in Section 3 are satisfied by any of the (G)ARCH models used in these experiments.

Lemma 4 *A GARCH(1,1) process, of the form* $y_t = \gamma\ y_{t-1} + \varepsilon_t$, $\varepsilon_t = z_t h_t$ *and* $h_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta h_{t-1}^2$, *where* $z_t \sim iid$, *has finite 8th moment if* $\beta^4 + 4\beta^3 \alpha_1 E\left(z_{t-1}^2\right) + 6\beta^2 \alpha_1^2 E\left(z_{t-1}^4\right) + 4\beta \alpha_1^3 E\left(z_{t-1}^6\right) + \alpha_1^4 E\left(z_{t-1}^8\right) < 1$ *and finite 16th moment if*

$$\begin{aligned} &\beta^8 + 8\beta^7 \alpha_1 E\left(z_{t-1}^2\right) + 28\beta^6 \alpha_1^2 E\left(z_{t-1}^4\right) + 56\beta^5 \alpha_1^3 E\left(z_{t-1}^6\right) + 70\beta^4 \alpha_1^4 E\left(z_{t-1}^8\right) \\ &+56\beta^3 \alpha_1^5 E\left(z_{t-1}^{10}\right) + 28\beta^2 \alpha_1^6 E\left(z_{t-1}^{12}\right) + 8\beta \alpha_1^7 E\left(z_{t-1}^{14}\right) + \alpha_1^8 E\left(z_{t-1}^{16}\right) < 1 \end{aligned}$$

Proof. Appendix. ■

In order to check this condition for the simulated (G)ARCH processes it is required to draw on the moments of the standard normal-, $\chi^2(2)$- and $t(5)$-distributed deviates. It is well known that only the first 4 moments of the $t(5)$ distribution are finite (Abramowitz and Stegun, 1972) and it is hence apparent that the (G)ARCH models simulated with $z_t \sim t(5)$ do not comply with the imposed moment restrictions. While higher moments for the $\chi^2(2)$ distribution exist it turns out that they do not guarantee finite higher moments for our GARCH processes. The combinations of parameter values for α_1 and β for which the 4th, 8th and 16th moments of a GARCH(1,1) process (with $z_t \sim N(0,1)$) exist are illustrated in Figure 1.

Lemma 5 *The GARCH3 process has finite 16th moments. The ARCH2 and GARCH2 process with* $N(0,1)$ *innovations have finite 8th moments but not finite 16th moments. The ARCH1 and GARCH1 processes with* $N(0,1)$ *innovations and all ARCH and GARCH processes with* $t(5)$ *and* $\chi^2(2)$ *innovations do not have finite 8th or higher moments.*

Proof. *Appendix.* ■

The location of the three GARCH(1,1) processes used in experiments are indicated by arrows. When GARCH(1,1) process are estimated using high-frequency financial data one often finds parameter estimates which are indeed closest to those of the *GARCH3* process,

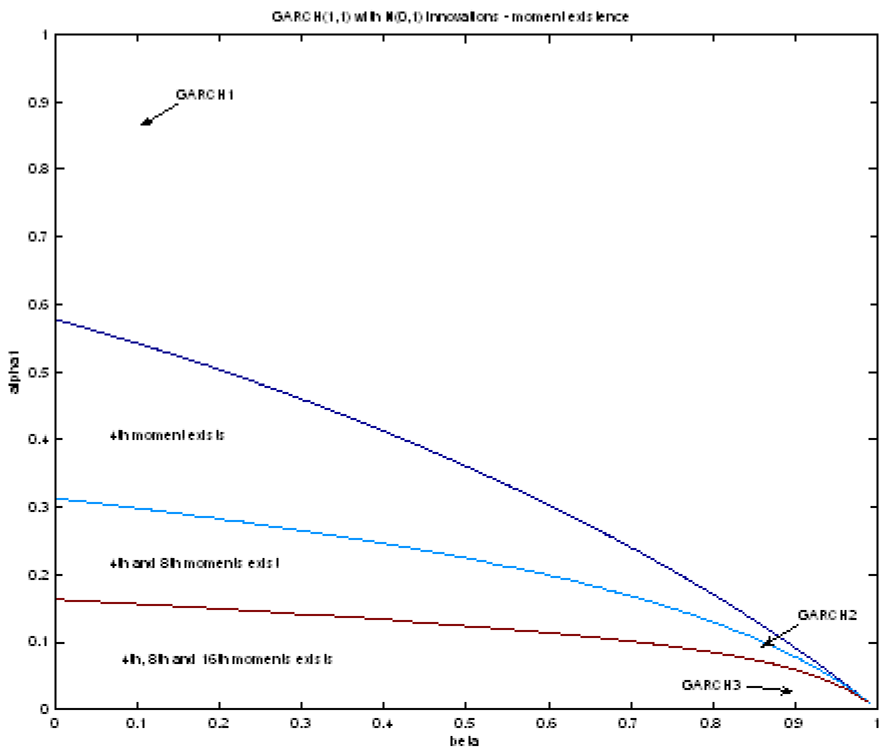

Figure 1: Illustrating moment existence for parameter combinations of a GARCH(1,1). The three models used in the simulation exercises are indicated by arrows.

which is the one that meets all moment requirements. The fact that the other DGPs fail to provide the required moment conditions is due to the use of third order cross products in the V23 test. The moments required for the theoretical results could be somewhat reduced if one was to use second order cross products only. The likely result of this, in practice however, would be a loss of power. While the theoretical results provided in this paper do not support the application of the fixed-design bootstrap to all the GARCH processes it is instructive to use all DGPs in the simulation design for at least three reasons. *First*, and most important, it is very difficult to establish the existence of moments in empirical data. Similarly, it is not straightforward to determine the distribution of a particular innovation process particularly in the light of uncertainty in choice of model specification. In this context, the development of a testing technique whose desirable properties are retained despite the fact that the required moment conditions are not met is of significant practical importance. For example, the parameter values used to generate the GARCH2 process are also often found in empirical applications and it will therefore be instructive to examine the performance of the test in these circumstances. *Second*, the skewed and leptokurtotic innovation processes were included, in a different context by Godfrey and Orme (2001, 2004), who established that the wild bootstrap has the potential to be robust to the types of non-normality introduced by these two innovation processes. In this regard the results reported here will provide further directly comparable evidence.

5 Simulation Results

In all the simulation results to follow the asymptotic V23 test is denoted simply as V23, the test based on the robust regression is denoted, V23hc, and the wild-bootstrap version of the robust-regression test is denoted V23wb, where R (F) indicates that the recursive (fixed design) algorithm has been applied. The size simulation results for normally distributed innovations are reported Table 1. A number of conclusions are immediately apparent.

1. The asymptotic V23 test rejects the correct null hypothesis far too frequently when there is heteroskedasticity in the data. The size distortions are especially dramatic

for those processes which have a strong ARCH component (ARCH1, ARCH2 and GARCH1). For the GARCH3 process the V23 test is indeed conservative.

2. The heteroskedastic-robust-regression test reduces substantially the size distortion of the simple V23 test when the latter is too liberal. The V23hc, however, does tend to be overly conservative, when the nominal size is small ($\alpha = 0.01$ or $\alpha = 0.05$). This is the case for all simulated processes although, as expected, the size distortion diminishes with increasing sample size. A similar result has been obtained by Godfrey and Orme (2004) in the context of testing several linear restrictions,

3. The wild-bootstrap version of the test, on the other hand, has the correct empirical size in all cases. These results suggest that the bootstrap can deal quite comfortably with heteroskedasticity of an autoregressive-conditional type. It should be noted that this conclusion is independent on whether the required moment conditions are met or not.

In general, it appears that simple robust-regression approach does offer potential benefits in correcting the size distortion suffered by tests of nonlinearity in the presence of ARCH. The benefit can be substantially enhanced by employing the wild bootstrap to determine the significance of the test statistic, particularly in small samples.

The power results for the four test statistics are reported in Table 2. Perhaps the most striking result is that the NAR process is not detected particularly reliably by any version of the test. This is not surprising as the NAR process is known to be notoriously difficult to detect. For the other data generating processes, especially the SAR, BILINAR and LSTAR, all the tests seem to have acceptable power. Furthermore the empirical power reported here is comparable to results reported in other studies (Lee *et al.,* 1993, Dahl, 1999).

From a purely practical point of view, it seems that using the robust-regression versions of the test when the data are not in fact heteroskedastic does not appear to result in significant decrease in power. There is, however, one caveat to this statement. It appears that for the BILIN class of models (BILIN and BILINAR) the power of the robust approaches is

		RNDN	AR1	ARCH1	ARCH2	GARCH1	GARCH2	GARCH3
					T = 50			
V23	0.01	0.005	0.006	0.179	0.089	0.198	0.012	0.006
	0.05	0.042	0.032	0.308	0.190	0.323	0.048	0.032
	0.10	0.088	0.072	0.389	0.271	0.405	0.097	0.077
V23wb (R)	0.01	0.013	0.015	0.018	0.016	0.015	0.015	0.011
	0.05	0.054	0.058	0.056	0.052	0.050	0.058	0.048
	0.10	0.104	0.103	0.106	0.102	0.106	0.107	0.095
V23wb (F)	0.01	0.013	0.010	0.021	0.018	0.012	0.011	0.001
	0.05	0.050	0.046	0.073	0.064	0.056	0.051	0.048
	0.10	0.108	0.089	0.129	0.116	0.106	0.106	0.094
V23hc	0.01	0.002	0.002	0.003	0.002	0.003	0.002	0.002
	0.05	0.026	0.025	0.038	0.034	0.037	0.027	0.025
	0.10	0.080	0.075	0.100	0.094	0.105	0.072	0.072
					T = 100			
V23	0.01	0.008	0.007	0.291	0.140	0.318	0.021	0.008
	0.05	0.042	0.038	0.426	0.265	0.450	0.075	0.046
	0.10	0.089	0.080	0.506	0.347	0.529	0.128	0.090
V23wb (R)	0.01	0.012	0.012	0.019	0.011	0.013	0.013	0.011
	0.05	0.049	0.053	0.059	0.054	0.057	0.053	0.051
	0.10	0.098	0.101	0.110	0.101	0.109	0.106	0.099
V23wb (F)	0.01	0.014	0.011	0.024	0.018	0.014	0.013	0.011
	0.05	0.055	0.048	0.069	0.066	0.059	0.049	0.049
	0.10	0.108	0.097	0.131	0.116	0.113	0.096	0.098
V23hc	0.01	0.003	0.004	0.005	0.004	0.005	0.004	0.003
	0.05	0.030	0.032	0.045	0.043	0.040	0.036	0.037
	0.10	0.083	0.083	0.113	0.107	0.108	0.086	0.084
					T = 200			
V23	0.01	0.007	0.009	0.419	0.205	0.152	0.032	0.007
	0.05	0.046	0.041	0.543	0.337	0.269	0.104	0.044
	0.10	0.092	0.084	0.618	0.425	0.357	0.171	0.090
V23wb (R)	0.01	0.012	0.013	0.015	0.014	0.014	0.013	0.011
	0.05	0.049	0.048	0.056	0056	0.053	0.051	0.053
	0.10	0.099	0.102	0.107	0.104	0.108	0.106	0.109
V23wb (F)	0.01	0.012	0.012	0.023	0.017	0.014	0.011	0.011
	0.05	0.053	0.052	0.074	0.063	0.056	0.053	0.053
	0.10	0.105	0.104	0.126	0.116	0.109	0.106	0.105
V23hc	0.01	0.004	0.004	0.006	0.004	0.004	0.004	0.003
	0.05	0.034	0.041	0.048	0.041	0.046	0.039	0.037
	0.10	0.085	0.092	0.118	0.104	0.099	0.094	0.084

Table 1: Size of the V23 test for nonlinearity in mean

		BILIN	TAR	SAR	NAR	BILINAR	LSTAR
					T = 50		
V23	0.01	0.578	0.094	0.205	0.015	0.589	0.360
	0.05	0.752	0.263	0.446	0.066	0.791	0.572
	0.10	0.825	0.396	0.591	0.122	0.870	0.681
V23wb (R)	0.01	0.050	0.173	0.318	0.022	0.123	0.124
	0.05	0.166	0.367	0.557	0.080	0.318	0.328
	0.10	0.284	0.482	0.690	0.139	0.457	0.471
V23wb (F)	0.01	0.053	0.175	0.305	0.023	0.127	0.153
	0.05	0.180	0.361	0.538	0.082	0.311	0.349
	0.10	0.296	0.480	0.675	0.146	0.464	0.510
V23hc	0.01	0.002	0.059	0.128	0.005	0.009	0.015
	0.05	0.059	0.261	0.430	0.045	0.131	0.175
	0.10	0.184	0.427	0.627	0.113	0.318	0.375
					T = 100		
V23	0.01	0.939	0.224	0.562	0.028	0.957	0.814
	0.05	0.980	0.460	0.801	0.101	0.986	0.916
	0.10	0.991	0.593	0.883	0.179	0.995	0.951
V23wb (R)	0.01	0.249	0.353	0.655	0.037	0.461	0.612
	0.05	0.493	0.587	0.850	0.118	0.715	0.825
	0.10	0.633	0.697	0.919	0.187	0.830	0.905
V23wb (F)	0.01	0.262	0.381	0.673	0.040	0.447	0.631
	0.05	0.502	0.603	0.866	0.116	0.702	0.833
	0.10	0.641	0.718	0.934	0.190	0.823	0.907
V23hc	0.01	0.071	0.210	0.491	0.012	0.169	0.353
	0.05	0.345	0.507	0.803	0.082	0.548	0.732
	0.10	0.569	0.662	0.905	0.166	0.750	0.865
					T = 200		
V23	0.01	0.997	0.525	0.928	0.047	1.00	0.994
	0.05	1.00	0.756	0.984	0.157	1.00	0.999
	0.10	1.00	0.845	0.995	0.254	1.00	1.00
V23wb (R)	0.01	0.604	0.697	0.943	0.071	0.854	0.981
	0.05	0.786	0.859	0.986	0.189	0.957	0.996
	0.10	0.859	0.908	0.995	0.285	0.974	0.998
V23wb (F)	0.01	0.626	0.690	0.952	0.079	0.863	0.983
	0.05	0.787	0.849	0.989	0.188	0.950	0.997
	0.10	0.858	0.910	0.997	0.284	0.974	0.999
v23hc	0.01	0.429	0.557	0.891	0.003	0.693	0.953
	0.05	0.737	0.807	0.979	0.147	0.916	0.994
	0.10	0.861	0.889	0.994	0.253	0.968	0.998

Table 2: Power of the V23 test for nonlinearity in mean

significantly less than for the test based on the asymptotic distribution. This power leakage is less noticeable for the wild bootstrap versions of the test and further decreases as the sample size increases to 200.

Finally in terms of the relative merits of the V23hc and V23wb tests, there appears little to choose between them when the nominal size is set at 10%. There does appear to be some difference between the tests when the nominal size is reduced. Here the conservative nature of the V23hc procedure, hinted at in the discussion of the size results, seems to manifest itself in a significant loss of power. If anything, therefore, the power results reinforce the conclusion reached previously that the bootstrap implementation of the robust-regression test is to be preferred. This conclusion ignores any computational considerations as bootstrapping of the heteroskedasticity-robust auxiliary regression is far more demanding in a computational sense then merely using the asymptotic distribution.

The Tables 3 and 4 illustrate how V23hc and V23wb fare when innovations are not normally distributed. As discussed previously, it is not necessary to assume normality in order to apply these tests and it is interesting to investigate their empirical properties when residuals are either leptokurtotic or skewed[5]. A comparison of the results in Table 1 and those in Table 3 indicate that the tendency of the robust regression approach to be conservative is slightly enhanced when $t\,(5)$ random deviates are used in the simulation. The V23wb, on the other hand, appears to be unaffected by the use of leptokurtotic errors in all simulated sample sizes. On balance it appears as if the recursive wild bootstrap test fares better for the models applied in this paper. The effects of skewed residuals are as expected with the performance of the robust regression V23 test being unaltered. This result is consistent with the observation that the absence of skewness is not a condition for the validity of the robust regression approach. As far as the wild bootstrap is concerned, the results reported by Godfrey and Orme (2001) are confirmed. They report that the particular version of the wild bootstrap applied in this paper is robust to skewed residuals in small samples. As sample sizes increase, however, a significant size distortion appears for the recursive wild bootstrap

[5] It should, however, be recalled that all but the GARCH3 process with normal innovations were shown to violate the process assumptions A.

		RNDN	AR1	ARCH1	ARCH2	GAR1	GAR2	GAR3
		T = 50						
V23	0.01	0.013	0.010	0.176	0.105	0.049	0.018	0.012
	0.05	0.046	0.035	0.286	0.197	0.128	0.058	0.040
	0.10	0.095	0.071	0.365	0.270	0.200	0.108	0.077
V23wb (R)	0.01	0.016	0.013	0.021	0.015	0.014	0.012	0.012
	0.05	0.057	0.057	0.062	0.059	0.054	0.054	0.050
	0.10	0.109	0.106	0.111	0.110	0.104	0.105	0.101
V23wb (F)	0.01	0.013	0.012	0.023	0.020	0.012	0.013	0.014
	0.05	0.050	0.053	0.076	0.068	0.056	0.054	0.050
	0.10	0.108	0.094	0.127	0.123	0.106	0.100	0.098
V23hc	0.01	0.001	0.001	0.002	0.003	0.002	0.001	0.001
	0.05	0.021	0.021	0.030	0.026	0.032	0.021	0.021
	0.10	0.066	0.061	0.085	0.080	0.085	0.066	0.063
		T = 100						
V23	0.01	0.014	0.007	0.288	0.173	0.090	0.036	0.012
	0.05	0.052	0.032	0.412	0.281	0.192	0.092	0.051
	0.10	0.096	0.069	0.490	0.367	0.272	0.145	0.083
V23wb (R)	0.01	0.014	0.012	0.016	0.016	0.011	0.013	0.014
	0.05	0.057	0.048	0.054	0.053	0.050	0.054	0.053
	0.10	0.107	0.096	0.105	0.108	0.102	0.102	0.103
V23wb (F)	0.01	0.014	0.010	0.025	0.022	0.014	0.012	0.013
	0.05	0.055	0.048	0.074	0.064	0.059	0.053	0.054
	0.10	0.108	0.092	0.126	0.117	0.113	0.101	0.102
V23hc	0.01	0.003	0.002	0.007	0.002	0.004	0.002	0.002
	0.05	0.025	0.021	0.036	0.028	0.042	0.028	0.024
	0.10	0.068	0.060	0.097	0.086	0.097	0.074	0.068
		T = 200						
V23	0.01	0.014	0.008	0.412	0.272	0.151	0.053	0.016
	0.05	0.045	0.035	0.537	0.386	0.276	0.134	0.057
	0.10	0.094	0.070	0.617	0.470	0.366	0.202	0.106
V23wb (R)	0.01	0.014	0.013	0.021	0.014	0.014	0.012	0.014
	0.05	0.057	0.057	0.062	0.055	0.054	0.054	0.056
	0.10	0.109	0.106	0.111	0.107	0.104	0.105	0.110
V23wb (F)	0.01	0.012	0.011	0.039	0.021	0.014	0.014	0.011
	0.05	0.053	0.052	0.103	0.065	0.056	0.048	0.054
	0.10	0.105	0.105	0.166	0.119	0.109	0.099	0.107
V23hc	0.01	0.002	0.002	0.006	0.004	0.005	0.002	0.002
	0.05	0.023	0.025	0.042	0.039	0.038	0.026	0.030
	0.10	0.066	0.068	0.108	0.098	0.100	0.075	0.077

Table 3: Size of the V23 test for nonlinearity in mean when residuals are t(5) distributed.

		RNDN	AR1	ARCH1	ARCH2	GAR1	GAR2	GAR3
				T = 50				
V23	0.01	0.012	0.009	0.142	0.816	0.048	0.013	0.001
	0.05	0.044	0.035	0.276	0.172	0.124	0.047	0.039
	0.10	0.083	0.070	0.366	0.254	0.198	0.093	0.073
V23wb (R)	0.01	0.014	0.017	0.027	0.024	0.011	0.020	0.022
	0.05	0.057	0.060	0.089	0.079	0.050	0.076	0.072
	0.10	0.107	0.119	0.148	0.134	0.102	0.129	0.125
V23wb (F)	0.01	0.011	0.010	0.021	0.017	0.012	0.012	0.013
	0.05	0.055	0.047	0.082	0.067	0.056	0.055	0.053
	0.10	0.106	0.098	0.150	0.128	0.106	0.109	0.101
V23hc	0.01	0.003	0.003	0.004	0.002	0.002	0.002	0.003
	0.05	0.029	0.025	0.045	0.035	0.032	0.028	0.025
	0.10	0.071	0.068	0.119	0.099	0.088	0.082	0.070
				T = 100				
V23	0.01	0.017	0.008	0.276	0.147	0.097	0.022	0.001
	0.05	0.056	0.029	0.411	0.272	0.198	0.076	0.038
	0.10	0.101	0.062	0.503	0.375	0.280	0.133	0.073
V23wb (R)	0.01	0.013	0.018	0.035	0.036	0.012	0.026	0.025
	0.05	0.054	0.060	0.098	0.089	0.054	0.073	0.069
	0.10	0.109	0.115	0.153	0.143	0.105	0.131	0.118
V23wb (F)	0.01	0.012	0.012	0.032	0.022	0.014	0.017	0.013
	0.05	0.053	0.047	0.095	0.075	0.059	0.058	0.053
	0.10	0.100	0.099	0.158	0.132	0.113	0.110	0.100
V23hc	0.01	0.005	0.003	0.010	0.008	0.004	0.006	0.004
	0.05	0.034	0.029	0.058	0.053	0.037	0.039	0.031
	0.10	0.077	0.075	0.130	0.114	0.097	0.089	0.076
				T = 200				
V23	0.01	0.014	0.007	0.423	0.259	0.159	0.038	0.013
	0.05	0.048	0.038	0.563	0.406	0.278	0.112	0.052
	0.10	0.088	0.073	0.634	0.498	0.364	0.190	0.103
V23wb (R)	0.01	0.014	0.023	0.036	0.034	0.014	0.026	0.023
	0.05	0.057	0.065	0.096	0.088	0.053	0.084	0.071
	0.10	0.109	0.118	0.155	0.137	0.108	0.140	0.121
V23wb (F)	0.01	0.014	0.012	0.039	0.026	0.014	0.021	0.015
	0.05	0.054	0.054	0.103	0.087	0.056	0.064	0.060
	0.10	0.108	0.102	0.166	0.145	0.109	0.117	0.110
V23hc	0.01	0.004	0.005	0.015	0.015	0.004	0.007	0.006
	0.05	0.031	0.037	0.073	0.066	0.041	0.044	0.040
	0.10	0.070	0.087	0.144	0.126	0.100	0.095	0.093

Table 4: Size of the V23 test for nonlinearity in mean when residuals are CHI square distributed with 2 degrees of freedom.

with the fixed design wild bootstrap showing significantly better size properties.

6 Empirical illustration

In order to illustrate the practical implementation of the robust tests for nonlinearity described in the paper, two data sets used by Lee *et al.* (1993) – LWG hereafter – are used, namely, the Japanese Yen/US Dollar exchange rate (monthly observations, 1974:1-1990:7) and the US three month Treasury bill interest rate (monthly observations, 1959:1-1990:7). They assumed that the residuals of the respective models under the null hypothesis were homoskedastic, although they recognized the implications for the size of the test should this assumption be violated. The results reported by LWG are now revisited and subjected to the three versions of the V23 test.

Turning first to the Yen / US$ exchange rate, the residuals from an AR(1) regression on the continuously-compounded returns are displayed in Figure 2. In testing these data for nonlinearity in mean LWG found that a number of tests failed to reject the null hypothesis of linearity. These included the Neural Network test (White, 1989), the RESET test (Ramsey, 1969), the McLeod-Li test (McLeod and Li, 1983) and the BDS test (Brock *et al.*, 1996). Only one test, the Bispectrum test (Hinich, 1982) rejected the null hypothesis of a linear AR(1) model for the exchange rate returns.

Since there is no clear indication of heteroskedasticity, at least in terms of volatility clustering, in the data, it is expected that the robust and non-robust versions of the V23 test come to the same conclusion. This indeed turns out to be the case as all three versions of the test fail to reject the null hypothesis of an AR(1) model. The p-values of the test statistics are as follows: V23 – 0.747, V23hc – 0.554, V23wb – 0.638. It seems reasonably safe to conclude that the log returns of the Yen/US$ exchange rate are linear in mean.

The situation is slightly different for the US 3-month Treasury Bill rate. The SIC criterion chooses an AR(6) model of the interest rate changes as the linear model under the null hypothesis and the residuals from this regression are plotted in Figure 3. A visual inspection clearly suggests the presence of autoregressive conditional heteroskedasticity in the

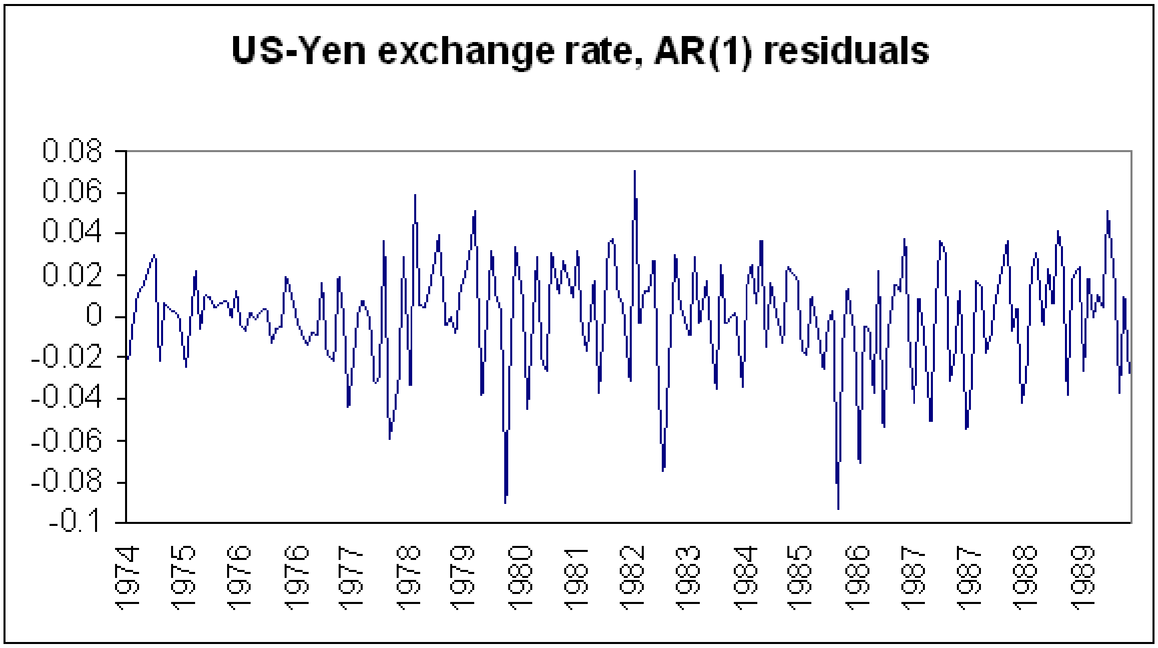

Figure 2: Residuals from an AR(1) model of the monthly changes of the logarithm of the US-YEn exchange rate.

interest rates. In particular the early 1980s are characterized by increased volatility, a fact which is widely attributed to the Federal Reserve's monetary experiment. The suspicion of heteroskedastic residuals is reinforced by the test results reported by LWG. The McLeod-Li (p-value of 0.00) and BDS (p-value of 0.00) tests, both of which are known to have power against ARCH, are highly significant.

Given the presence of heteroskedasticity, the results of the other non-robust tests for nonlinearity in mean reported by LWG, all of which indicate a solid rejection of linearity, are to be interpreted with extreme care. This note of caution is reinforced by the results of the V23 test in its various forms. The asymptotic V23 test records a p-value of 0.00 clearly in line with the results reported by LWG. The V23hc while not allowing rejection of the null at the 1%, as is the case for all the tests reported by LWG, is significant at the 5% level. The preferred wild-bootstrap version of the V23 test, however, records a p-value of 0.154, indicating that the null hypothesis of a linear specification for the mean cannot be rejected even at a 10% significance level. The application of a heteroskedasticity robust test for

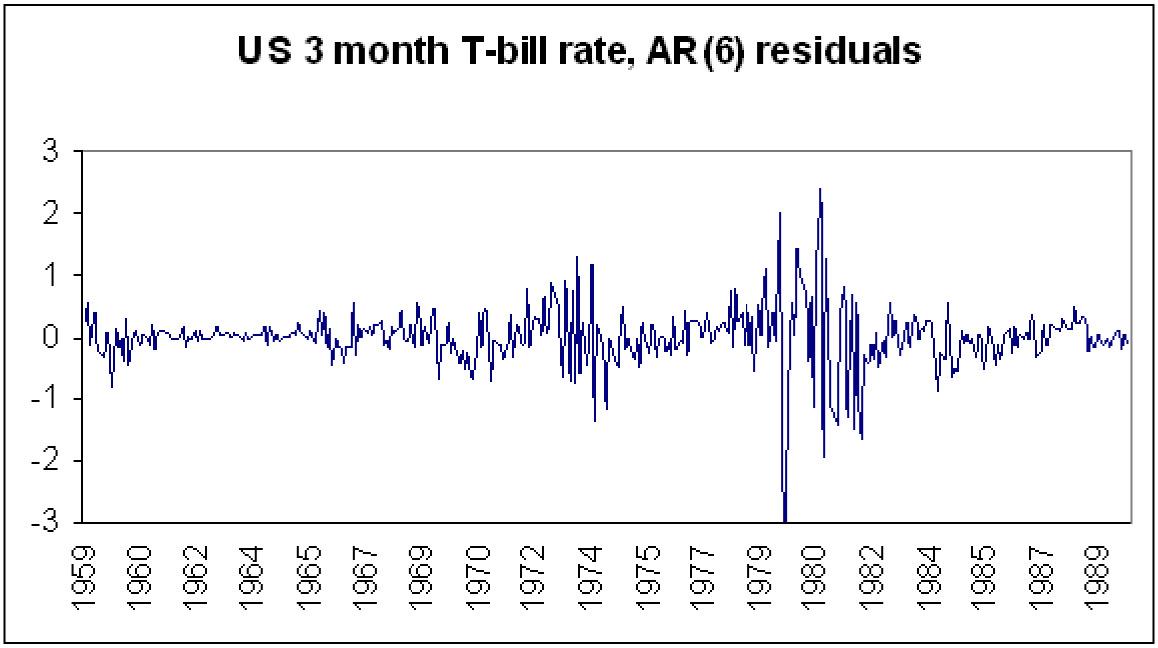

Figure 3: Residuals from an AR(6) model of the monthly changes of the US 3 month T-bill rate.

nonlinearity appears to be crucial in the context of this interest rate data set. What initially looked like a clear rejection of the null hypothesis of linearity is turned into a, at most, marginal rejection. The clear rejection of the null hypothesis, suggested by the non-adjusted tests, is not justified by the data.

7 Conclusion

This paper has addressed an important practical problem faced when testing for nonlinearity in mean in time-series data, namely that tests reject the null hypothesis of linearity too frequently when the data have non-constant variance. This is a particularly acute problem given the prevalence of autoregressive conditional heteroskedasticity in most economic and financial time series. A testing strategy based on the heteroskedastic-robust-auxiliary regression and the wild bootstrap is proposed and its empirical performance evaluated in this paper. Monte Carlo experiments verify that the approach has the potential to eliminate the observed size distortion and that this improvement comes without significant loss of power in most cases. The results indicate a slight preference for the wild bootstrap version of the heteroskedastic-consistent test. The results appeared to be fairly robust against the failure of moment existence requirements, which is a convenient finding, as the moment requirements for the validity of the V23 test are rather strict. Two empirical examples of the testing strategies are provided which emphasize the need for caution in interpreting the results of nonlinearity tests which are not robust to the presence of heteroskedasticity.

In order to come to these conclusions, a number of theoretical results had to be established. First the applicability of the V23 test to GARCH type processes was investigated and necessary process assumptions established. It was then shown that these assumptions do not allow for the particular non-normality introduced by the chosen t- and χ^2- distribution. Last, it was shown that, given an asymptotic χ^2- distribution of the V23 test, the fixed-design wild bootstrap can consistently replicate this distribution.

Several issues warrant further investigation. In this paper attention has been focussed on a single test for nonlinearity. Further simulation with other suitable tests will provide more evidence on the efficacy of both these robust testing strategies. Finally, it is important to note that the heteroskedasticity examined in this paper is limited to the GARCH class. Self-evidently the robustness of the tests to GARCH cannot automatically be taken to extend to other types of heteroskedasticity.

A Simulated DGPs

All but the ARCH and GARCH data generating processes in this study have been used before in either Lee *et al.* (1993) or Teräsvirta *et al.* (1994).

Autoregressive model (AR1):

$$y_t = 0.6\ y_{t-1} + \varepsilon_t$$

Bilinear model (BILIN):

$$y_t = 0.7\ y_{t-1}\varepsilon_{t-2} + \varepsilon_t$$

Threshold autoregressive model (TAR):

$$\begin{aligned} y_t &= 0.9\ y_{t-1} + \varepsilon_t\ for\ |y_{t-1}| \leq 1 \\ &= -0.3\ y_{t-1} + \varepsilon_t\ for\ |y_{t-1}| > 1 \end{aligned}$$

Sign autoregressive model (SAR):

$$y_t = sgn(y_{t-1}) + \varepsilon_t$$

where $sgn(x) = 1$ for all $x > 1$, $sgn(x) = 0$ for $x = 0$ and $sgn(x) = -1$ for all $x < 1$.

Nonlinear autoregressive (NAR):

$$y_t = (0.7\ |y_{t-1}|)/(|y_{t-1}| + 2) + \varepsilon_t$$

Bilinear autoregressive model (BILINAR):

$$y_t = 0.4\ y_{t-1} - 0.3\ y_{t-2} + 0.5\ y_{t-1}\varepsilon_{t-1} + \varepsilon_t$$

Logistic smooth transition autoregression (LSTAR):

$$\begin{aligned} y_t &= (0.0 + 0.02F_t) + (1.8 - 0.9F_t)\ y_{t-1} \\ &\quad +(-1.06 + 0.795F_t)\ y_{t-2} + \varepsilon_t \\ \text{where } F_t &= \left[1 + \exp\left(100\left(y_{t-1} - 0.02\right)\right)\right]^{-1} \\ \text{and } \varepsilon_t &\sim N(0, 0.02^2) \end{aligned}$$

ARCH :

$$\begin{aligned} y_t &= 0.5\ y_{t-1} + \varepsilon_t, \text{ where } \varepsilon_t \sim N(0, h_t) \\ \text{ARCH1} &: h_t = 1 + 0.8\ \varepsilon_{t-1}^2 \\ \text{ARCH2} &: h_t = 1 + 0.3\ \varepsilon_{t-1}^2 \end{aligned}$$

GARCH :

$$\begin{aligned} y_t &= 0.5\ y_{t-1} + \varepsilon_t, \text{ where } \varepsilon_t \sim N(0, h_t) \\ \text{GARCH1} &: h_t = 1 + 0.85\ \varepsilon_{t-1}^2 + 0.1\ h_{t-1} \\ \text{GARCH2} &: h_t = 1 + 0.1\ \varepsilon_{t-1}^2 + 0.85\ h_{t-1} \\ \text{GARCH3} &: h_t = 1 + 0.02\ \varepsilon_{t-1}^2 + 0.90\ h_{t-1} \end{aligned}$$

If not stated otherwise the error term ε_t was drawn from a standard normal distribution.

B Technical Results

The following Lemma will be useful throughout this Appendix.

Lemma 6 *Given the DGP $\phi(L)\, y_t = \varepsilon_t$ where the polynomial order is known and all roots outside the unit circle, the $(p \times 1)$ vector $\mathbf{y}_{t-1} = (y_{t-1}, \ldots, y_{t-p})'$ can be represented as follows:*

$$\mathbf{y}_{t-1} = \sum_{j=1}^{\infty} \mathbf{b}_j \varepsilon_{t-j}$$

where $\mathbf{b}_j = \left(\psi_{j-1}, \ldots, \psi_{j-p}\right)'$. Further note that the second and third order cross products of y_t can be written as follows:

$$\begin{aligned}
y_{t-k} y_{t-l} &= \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-k-j} \sum_{n=0}^{\infty} \psi_n \varepsilon_{t-l-n} = \sum_{j=0}^{\infty} \sum_{n=0}^{\infty} \psi_j \psi_n \varepsilon_{t-k-j} \varepsilon_{t-l-n} \\
y_{t-k} y_{t-l} y_{t-m} &= \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-k-i} \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-l-j} \sum_{n=0}^{\infty} \psi_n \varepsilon_{t-m-n} \\
&= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{n=0}^{\infty} \psi_i \psi_j \psi_n \varepsilon_{t-k-i} \varepsilon_{t-l-j} \varepsilon_{t-m-n}.
\end{aligned}$$

Proof. The proof of the first part is from Gonçalves and Kilian (2004). The process under consideration is

$$\beta(L)\, y_t = \varepsilon_t \tag{13}$$

where the autoregression coefficient lag polynomial of known order p is $\beta(L) = 1 - \beta_1 L - \beta_2 L^2 - \ldots - \beta_p L^p$, assuming that β_p is non-zero and all roots outside the unit circle. Further process assumptions are those used by GK in their set of assumptions A, which are a subset of the assumptions made here. The assumptions are general enough to allow for a GARCH(p,q) error process but do exclude some more complicated asymmetric GARCH-type processes. Given stationarity the process in (13) can be represented by an infinite order MA process

$$y_t = \beta^{-1}(L)\, \varepsilon_t = \psi(L)\, \varepsilon_t \tag{14}$$

where $\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$. A further necessary piece of notation is $\mathbf{b}_j = \left(\psi_{j-1}, \ldots, \psi_{j-p}\right)'$, noting that $\psi_0 = 1$ and $\psi_j = 0$ for all $j < 0$. Define the $(p \times 1)$ vector $\mathbf{y}_{t-1} = (y_{t-1}, \ldots, y_{t-p})'$ and note that $\mathbf{y}_{t-1}$ can be restated as

$$\mathbf{y}_{t-1} = (\psi(L)\varepsilon_{t-1}, \ldots, \psi(L)\varepsilon_{t-p})' = \left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-1-j}, \ldots, \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-p-j}\right)' = \sum_{j=1}^{\infty} \mathbf{b}_j \varepsilon_{t-j}.$$

The second part of the Lemma follows immediately. ■

The next Lemma is from Halunga (2005, Lemma 2) and will be used repeatedly below.

Lemma 7 *Let* $\{Z_{t,i}\}_{t=1}^{T}$ *be a sequence of random variables, for fixed* i*, such that (i)* $E|Z_{t,i}|$ *is uniformly bounded, and (ii)* $\overline{Z}_{T,i} = T^{-1}\sum_{t=1}^{T} Z_{t,i} \xrightarrow{p} 0$*. Define* $\overline{W}_T = T^{-1}\sum_{t=1}^{T} W_t = \sum_{i=1}^{\infty} h_i \overline{Z}_{T,i}$*, where* $\sum_{i=1}^{\infty}|h_i| < \infty$ *and* $W_t = \sum_{i=1}^{\infty} h_i Z_{t,i}$*. Then (a)* $\overline{W}_T \xrightarrow{p} 0$*; and (b)* $sup_t E|W_t| \leq \Delta < \infty$*.*

Proof. Lemma 7. Halunga (2005) ■

The following Lemmas are used in the proof to Lemma 2.

Lemma 8 *Given Assumptions A,* $\left\{\mathbf{y}_{t-1}'\mathbf{y}_{t-1}\right\}$ *satisfies the UWLLN and UC assumption.*

Proof. Gonçalves and Kilian (2004, Theorem 3.1) establish that $T^{-1}\sum_{t=1}^{T} \mathbf{y}_{t-1}'\mathbf{y}_{t-1} \equiv A_{1T} \xrightarrow{p} A$, where $A = \sigma^2 \sum_{j=1}^{\infty} \mathbf{b}_j \mathbf{b}_j'$. ■

Lemma 9 *Given Assumptions A,* $\left\{\mathbf{y}_{t-1}'\boldsymbol{\lambda}_t\right\}$ *satisfies the UWLLN and UC assumption.*

Proof. We need to show that $T^{-1}\sum_{t=1}^{T} \mathbf{y}_{t-1}'\boldsymbol{\lambda}_t - E\left(\mathbf{y}_{t-1}'\boldsymbol{\lambda}_t\right) \xrightarrow{p} 0$. The proof will follow the same lines of argument as that in Halunga (2005) and indeed make repeated use of Lemma 7. The vector $\boldsymbol{\lambda}_t$ contains 2nd and 3rd order cross products of the p elements in $\mathbf{y}_{t-1}$. $\mathbf{y}_{t-1}'\boldsymbol{\lambda}_t$ therefore contains 3rd and 4th order cross products.

Consider first a typical 3rd order product $y_{t-k}y_{t-l}y_{t-m}$ with $1 \leq k \leq l \leq m \leq p$. It will be demonstrated that (i) $E(y_{t-k}y_{t-l}y_{t-m}) = B_3 = \sigma^3 \sum_{i=0}^{\infty}\sum_{n=0}^{\infty} \psi_i^2 \psi_n \tau_{k,l,m}$ if $k = l$ and $B_3 = 0$ otherwise. Further we will show that that (ii) $B_{3T} \equiv T^{-1}\sum_{t=1}^{T} y_{t-k}y_{t-l}y_{t-m} \xrightarrow{p} B_3$.To show (i): $B_3 = E[\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{n=0}^{\infty} \psi_i\psi_j\psi_n \varepsilon_{t-k-i}\varepsilon_{t-l-j}\varepsilon_{t-m-n}]$ from Lemma DGP.

$B_3 = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{n=0}^{\infty}\psi_i\psi_j\psi_n E\left[\varepsilon_{t-k-i}\varepsilon_{t-l-j}\varepsilon_{t-m-n}\right]$ and $B_3 = \sigma^3\sum_{i=0}^{\infty}\sum_{n=0}^{\infty}\psi_i^2\psi_n\tau_{k,l,m}$ follow immediately from assumption A8 and $B_3 = 0$ for $k \neq l$ follows from the mds property of ε_t. To show (ii) consider the difference

$$\begin{aligned} & \left(B_{3T} - B_3\right) \\ = \; & T^{-1}\sum_{t=1}^{T}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{n=0}^{\infty}\psi_i\psi_j\psi_n\left(\varepsilon_{t-k-i}\varepsilon_{t-l-j}\varepsilon_{t-m-n} - E\left[\varepsilon_{t-k-i}\varepsilon_{t-l-j}\varepsilon_{t-m-n}\right]\right). \end{aligned}$$

Let $\boldsymbol{\lambda} \in R^p$ and $\boldsymbol{\lambda}'\boldsymbol{\lambda}$ =1, and define $h_i = \boldsymbol{\lambda}'\psi_i$ as well as

$$\begin{aligned} \overline{Z}_T &= \boldsymbol{\lambda}'\left(B_{3T} - B_3\right)\boldsymbol{\lambda} \\ &= T^{-1}\sum_{t=1}^{T}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{n=0}^{\infty}h_ih_jh_n\left(\varepsilon_{t-k-i}\varepsilon_{t-l-j}\varepsilon_{t-m-n} - E\left[\varepsilon_{t-k-i}\varepsilon_{t-l-j}\varepsilon_{t-m-n}\right]\right). \end{aligned}$$

As $\sum_{n=0}^{\infty}|\psi_i| < \infty$, $\sum_{n=0}^{\infty}|h_i| \leq \Delta < \infty$ follows. In order to establish (ii) $\overline{Z}_T \xrightarrow{p} 0$ is required. The trick is to redefine the processes such that a core process without any of the parameter series remains. The process assumptions will then allow the repeated application of Lemma 7. Define $\overline{Z}_T = \sum_{i=0}^{\infty}h_i\overline{Z}_{T,i}$, $\overline{Z}_{T,i} = \sum_{j=0}^{\infty}h_j\overline{Z}_{T,i,j}$ and $\overline{Z}_{T,i,j} = \sum_{n=0}^{\infty}h_n\overline{Z}_{T,i,j,n}$ where

$$\overline{Z}_{T,i,j,n} = T^{-1}\sum_{t=1}^{T}\left(\varepsilon_{t-k}\varepsilon_{t-l}\varepsilon_{t-m} - E\left[\varepsilon_{t-k}\varepsilon_{t-l}\varepsilon_{t-m}\right]\right).$$

At this stage we can differentiate the two cases (a) $k < l$ and (b) $k = l$. First consider case (a): From A4 we know that $E\left[\varepsilon_{t-k}\varepsilon_{t-l}\varepsilon_{t-m}\right] = 0$ such that $\overline{Z}_{T,i,j,n} = T^{-1}\sum_{t=1}^{T}\varepsilon_{t-k}\varepsilon_{t-l}\varepsilon_{t-m}$. From A1 we know that $\left\{\varepsilon_{t-k}\varepsilon_{t-l}\varepsilon_{t-m}\left|\mathcal{F}_{t-k}\right.\right\}$ is a mds and hence, provided $E\left|\varepsilon_{t-k}\varepsilon_{t-l}\varepsilon_{t-m}\right|^{1+\delta}$ is uniformly bounded for any $\delta > 0$, the application of Andrews LLN will establish $\overline{Z}_{T,i,j,n} \xrightarrow{p} 0$. Application of the Cauchy-Schwartz inequality yields

$$E\left|\varepsilon_{t-k}\varepsilon_{t-l}\varepsilon_{t-m}\right|^{1+\delta} \leq \left(E\left|\varepsilon_{t-k}\varepsilon_{t-l}\right|^{2+2\delta}E\left|\varepsilon_{t-m}\right|^{2+2\delta}\right)^{1/2}.$$

This implies that boundedness of $E\left|\varepsilon_{t-k}\varepsilon_{t-l}\varepsilon_{t-m}\right|^{1+\delta}$ requires boundedness of $E\left|\varepsilon_{t-k}\varepsilon_{t-l}\right|^{2+\widetilde{\delta}}$ and $E\left|\varepsilon_{t-m}\right|^{2+\widetilde{\delta}}$ for any $\widetilde{\delta} > 0$. The latter is guaranteed by A6. After another application of the Cauchy-Schwartz inequality it is apparent that the former is true provided $E\left|\varepsilon_{t-k}\right|^{4+\overline{\delta}}$ and $E\left|\varepsilon_{t-l}\right|^{4+\overline{\delta}}$ for any $\overline{\delta} > 0$ which is sanctioned by A6. Therefore $\overline{Z}_{T,i,j,n} \xrightarrow{p} 0$ is shown. This further establishes the assumption required for the application of Lemma AH and $\overline{Z}_{T,i,j} = \sum_{n=0}^{\infty}h_n\overline{Z}_{T,i,j,n} \xrightarrow{p} 0$ is therefore also shown. Two more applications of this Lemma establish that $\overline{Z}_{T,i} \xrightarrow{p} 0$ and $\overline{Z}_T \xrightarrow{p} 0$ and hence (ii)(a). To show (b) we A4 we know recall that $E\left[\varepsilon_{t-k}^2\varepsilon_{t-m}\right] = \sigma^3\tau_{k,k,m}$ such that

$$\begin{aligned}
\overline{Z}_{T,i,j,n} &= T^{-1}\textstyle\sum_{t=1}^T \left(\varepsilon_{t-k}^2\varepsilon_{t-m} - \sigma^3\tau_{k,k,m}\right) \\
&= T^{-1}\textstyle\sum_{t=1}^T \left(\varepsilon_{t-k}^2\varepsilon_{t-m} - E\left(\varepsilon_{t-k}^2\varepsilon_{t-m}\left|\mathcal{F}_{t-k-1}\right.\right)\right) \\
&\quad +T^{-1}\textstyle\sum_{t=1}^T \left(E\left(\varepsilon_{t-k}^2\varepsilon_{t-m}\left|\mathcal{F}_{t-k-1}\right.\right) - \sigma^3\tau_{k,k,m}\right) \\
&= T^{-1}\textstyle\sum_{t=1}^T A_1 + T^{-1}\sum_{t=1}^T A_2.
\end{aligned}$$

By definition A_1 is a mds and if $E\left|\varepsilon_{t-k}^2\varepsilon_{t-m} - E\left(\varepsilon_{t-k}^2\varepsilon_{t-m}\left|\mathcal{F}_{t-k-1}\right.\right)\right|^{1+\delta}$ is bounded Andrews LLN will ensure $T^{-1}\sum_{t=1}^T A_1 \xrightarrow{p} 0$. Note that by the c_r inequality

$$\begin{aligned}
E\left|\varepsilon_{t-k}^2\varepsilon_{t-m} - E\left(\varepsilon_{t-k}^2\varepsilon_{t-m}\left|\mathcal{F}_{t-k-1}\right.\right)\right|^{1+\delta} &\leq 2^\delta\left(E\left|\varepsilon_{t-k}^2\varepsilon_{t-m}\right|^{1+\delta} + E\left|E\left(\varepsilon_{t-k}^2\varepsilon_{t-m}\left|\mathcal{F}_{t-k-1}\right.\right)\right|^{1+\delta}\right) \\
&\leq 2^\delta\left(2\sup_{t\geq 1}E\left|\varepsilon_{t-k}^2\varepsilon_{t-m}\right|^{1+\delta}\right) = 2^{1+\delta}\sup_{t\geq 1}E\left|\varepsilon_{t-k}^2\varepsilon_{t-m}\right|^{1+\delta}
\end{aligned}$$

Another application of the Cauchy-Schwartz inequality yields

$$E\left|\varepsilon_{t-k}^2\varepsilon_{t-m} - E\left(\varepsilon_{t-k}^2\varepsilon_{t-m}\left|\mathcal{F}_{t-k-1}\right.\right)\right|^{1+\delta} \leq 2^{1+\delta}\sup_{t\geq 1}\left(\left(E\left|\varepsilon_{t-k}^4\right|^{1+\delta}E\left|\varepsilon_{t-m}^2\right|^{1+\delta}\right)^{1/2}\right)$$

which by A6 is bounded. This guarantees the application of the Andrews LLN and hence establishes $T^{-1}\sum_{t=1}^T A_1 \xrightarrow{p} 0$. Finally $T^{-1}\sum_{t=1}^T A_2 \xrightarrow{p} 0$ by A5, which establishes that $\overline{Z}_{T,i,j,n} \xrightarrow{p} 0$. Lemma AH then proves that $\overline{Z}_{T,i,j} \xrightarrow{p} 0$ and repeating the application of this Lemma two more times finally establishes $\overline{Z}_T \xrightarrow{p} 0$.

The proof for a fourth order cross product is completely analogous but requires one additional application of the Cauchy-Schwartz inequality and hence the existence of $E\left|\varepsilon_{t-k}^8\right|^{1+\delta}$, which is guaranteed by A6. ■

Lemma 10 *Given Assumptions A, $\{\boldsymbol{\lambda}_t'\varepsilon_t\varepsilon_t'\boldsymbol{\lambda}_t\}$ satisfies the UWLLN and UC assumption.*

Proof. The elements in the $(1\times q)$ vector $\boldsymbol{\lambda}_t'\varepsilon_t$ are of the following form: $\widetilde{\lambda}_{t-1}^{(k,l)} = \sum_{i=0}^\infty\sum_{j=0}^\infty \psi_i\psi_j\varepsilon_{t-k-i}\varepsilon_{t-l-j}\varepsilon_t$ for 2nd order cross products in $\boldsymbol{\lambda}_t$ and $\widetilde{\lambda}_{t-1}^{(k,l,m)} = \sum_{i=0}^\infty\sum_{j=0}^\infty \sum_{n=0}^\infty \psi_i\psi_j\psi_n\varepsilon_{t-k-i}\varepsilon_{t-l-j}\varepsilon_{t-m-n}\varepsilon_t$ for 3rd order elements in $\boldsymbol{\lambda}_t$. Elements in $\{\boldsymbol{\lambda}_t'\varepsilon_t\varepsilon_t'\boldsymbol{\lambda}_t\}$will then be cross products of $\widetilde{\lambda}_{t-1}^{(k,l)}$ and $\widetilde{\lambda}_{t-1}^{(k,l,m)}$. It is apparent that this will result in terms with cross products of the form $\{\varepsilon_t^2\widetilde{\varepsilon}_{t-1}\}$ where $\widetilde{\varepsilon}_{t-1}$ is a cross product of either order 4, 5 or 6. In order to show that $\{\boldsymbol{\lambda}_t'\varepsilon_t\varepsilon_t'\boldsymbol{\lambda}_t\}$satisfies a WLLN it will be demonstrated that (i)

$$\begin{aligned} E\left(\varepsilon_t^2 y_{t-k_1}y_{t-l_1}y_{t-m_1}y_{t-k_2}y_{t-l_2}y_{t-m_2}\right) &= B_8 \\ &= \sigma^8 \sum_{i_1=0}^{\infty}...\sum_{i_6=0}^{\infty}\psi_{i_1}\cdot...\cdot\psi_{i_6}\tau_{0,0,k_1,l_1,m_1,k_2,l_2,m_2}. \end{aligned}$$

Further we will show that that (ii) $B_{3T} \equiv T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 y_{t-k_1}y_{t-l_1}y_{t-m_1}y_{t-k_2}y_{t-l_2}y_{t-m_2} \xrightarrow{p} B_8$.

To show (i):

$$\begin{aligned} B_8 &= E[\varepsilon_t^2 \sum_{i_1=0}^{\infty}...\sum_{i_6=0}^{\infty}\psi_{i_1}\cdot...\cdot\psi_{i_6} \\ &\quad \cdot\varepsilon_{t-k_1-i_1}\varepsilon_{t-l_1-i_2}\varepsilon_{t-m_1-i_3}\varepsilon_{t-k_2-i_4}\varepsilon_{t-l_2-i_5}\varepsilon_{t-m_2-i_6}] \\ &= E[\varepsilon_t^2 \sum_{i_1=0}^{\infty}...\sum_{i_6=0}^{\infty}\psi_{i_1}\cdot...\cdot\psi_{i_6}\widetilde{\varepsilon}_{t-1} \\ &\quad \text{where} \\ \widetilde{\varepsilon}_{t-1} &= \varepsilon_{t-k_1-i_1}\varepsilon_{t-l_1-i_2}\varepsilon_{t-m_1-i_3}\varepsilon_{t-k_2-i_4}\varepsilon_{t-l_2-i_5}\varepsilon_{t-m_2-i_6} \end{aligned}$$

from Lemma DGP. $B_8 = \sum_{i_1=0}^{\infty}...\sum_{i_6=0}^{\infty}\psi_{i_1}\cdot...\cdot\psi_{i_6}E\left[\varepsilon_t^2\widetilde{\varepsilon}_{t-1}\right]$ and $B_8 = \sigma^8\sum_{i_1=0}^{\infty}...\sum_{i_6=0}^{\infty}\psi_{i_1}\cdot ...\cdot\psi_{i_6}\tau_{0,0,k_1,l_1,m_1,k_2,l_2,m_2}$ follow immediately from assumption A4. To show (ii) consider the difference

$$\begin{aligned} &(B_{8T}-B_8) \\ &= T^{-1}\sum_{t=1}^{T}\sum_{i_1=0}^{\infty}...\sum_{i_6=0}^{\infty}\psi_{i_1}\cdot...\cdot\psi_{i_6}\left(\varepsilon_t^2\widetilde{\varepsilon}_{t-1} - E\left[\varepsilon_t^2\widetilde{\varepsilon}_{t-1}\right]\right). \end{aligned}$$

Let $\boldsymbol{\lambda} \in R^p$ and $\boldsymbol{\lambda}'\boldsymbol{\lambda}$ =1, and define $h_i = \boldsymbol{\lambda}'\psi_i$ as well as

$$\begin{aligned} \overline{Z}_T &= \boldsymbol{\lambda}'\left(B_{3T}-B_3\right)\boldsymbol{\lambda} \\ &= T^{-1}\sum_{t=1}^{T}\sum_{i_1=0}^{\infty}...\sum_{i_6=0}^{\infty}h_{i_1}\cdot...\cdot h_{i_6}\left(\varepsilon_t^2\widetilde{\varepsilon}_{t-1} - E\left[\varepsilon_t^2\widetilde{\varepsilon}_{t-1}\right]\right). \end{aligned}$$

As $\sum_{n=0}^{\infty}|\psi_i| < \infty$, $\sum_{n=0}^{\infty}|h_i| \leq \Delta < \infty$ follows. In order to establish (ii) $\overline{Z}_T \xrightarrow{p} 0$ is required. The processes will be redefined such that a core process without any of the parameter series remains. The process assumptions will then allow the repeated application of Lemma 7. Define $\overline{Z}_T = \sum_{i_1=0}^{\infty}h_{i_1}\overline{Z}_{T,i_1}$ and $\overline{Z}_{T,i_n} = \sum_{j=0}^{\infty}h_{i_{n+1}}\overline{Z}_{T,i_{n+1}}$ for $n = 1, ..., 5$ where

$$\overline{Z}_{T,i_6} = T^{-1}\sum_{t=1}^{T}\left(\varepsilon_t^2\widetilde{\varepsilon}_{t-1} - E\left[\varepsilon_t^2\widetilde{\varepsilon}_{t-1}\right]\right).$$

To allow the application of Lemma 7 it is required to show that $\overline{Z}_{T,i_6} \xrightarrow{p} 0$. From A4 we recall that $E\left[\varepsilon_t^2 \widetilde{\varepsilon}_{t-1}\right] = \sigma^8 \tau_{0,0,k_1,l_1,m_1,k_2,l_2,m_2}$ such that

$$\begin{aligned}
\overline{Z}_{T,i_6} &= T^{-1} \textstyle\sum_{t=1}^{T} \left(\varepsilon_t^2 \widetilde{\varepsilon}_{t-1} - \sigma^8 \tau_{0,0,k_1,l_1,m_1,k_2,l_2,m_2}\right) \\
&= T^{-1} \textstyle\sum_{t=1}^{T} \left(\varepsilon_t^2 \widetilde{\varepsilon}_{t-1} - E\left(\varepsilon_t^2 \widetilde{\varepsilon}_{t-1} \left|\mathcal{F}_{t-1}\right.\right)\right) \\
&\quad + T^{-1} \textstyle\sum_{t=1}^{T} \left(E\left(\varepsilon_t^2 \widetilde{\varepsilon}_{t-1} \left|\mathcal{F}_{t-1}\right.\right) - \sigma^8 \tau_{0,0,k_1,l_1,m_1,k_2,l_2,m_2}\right) \\
&= T^{-1} \textstyle\sum_{t=1}^{T} A_1 + T^{-1} \textstyle\sum_{t=1}^{T} A_2.
\end{aligned}$$

By definition A_1 is a mds and if $E\left|\varepsilon_t^2 \widetilde{\varepsilon}_{t-1} - E\left(\varepsilon_t^2 \widetilde{\varepsilon}_{t-1} \left|\mathcal{F}_{t-1}\right.\right)\right|^{1+\delta}$ is bounded Andrews LLN will ensure $T^{-1} \sum_{t=1}^{T} A_1 \xrightarrow{p} 0$. Note that by the c_r inequality

$$\begin{aligned}
E\left|\varepsilon_t^2 \widetilde{\varepsilon}_{t-1} - E\left(\varepsilon_t^2 \widetilde{\varepsilon}_{t-1} \left|\mathcal{F}_{t-1}\right.\right)\right|^{1+\delta} &\leq 2^{\delta}\left(E\left|\varepsilon_t^2 \widetilde{\varepsilon}_{t-1}\right|^{1+\delta} + E\left|E\left(\varepsilon_t^2 \widetilde{\varepsilon}_{t-1} \left|\mathcal{F}_{t-1}\right.\right)\right|^{1+\delta}\right) \\
&\leq 2^{\delta}\left(2 \sup_{t\geq 1} E\left|\varepsilon_t^2 \widetilde{\varepsilon}_{t-1}\right|^{1+\delta}\right) = 2^{1+\delta} \sup_{t\geq 1} E\left|\varepsilon_t^2 \widetilde{\varepsilon}_{t-1}\right|^{1+\delta}
\end{aligned}$$

Another application of the Cauchy-Schwartz inequality yields

$$E\left|\varepsilon_t^2 \widetilde{\varepsilon}_{t-1} - E\left(\varepsilon_t^2 \widetilde{\varepsilon}_{t-1} \left|\mathcal{F}_{t-1}\right.\right)\right|^{1+\delta} \leq 2^{1+\delta} \sup_{t\geq 1}\left(\left(E\left|\varepsilon_t^4\right|^{1+\delta} E\left|\widetilde{\varepsilon}_{t-1}^2\right|^{1+\delta}\right)^{1/2}\right)$$

where we know that $E\left|\varepsilon_t^4\right|^{1+\delta}$ is bounded by A6. It remains to establish the boundedness of

$$\begin{aligned}
E\left|\widetilde{\varepsilon}_{t-1}^2\right|^{1+\delta} &= E\left|\varepsilon_{t-k_1}^2 \varepsilon_{t-l_1}^2 \varepsilon_{t-m_1}^2 \varepsilon_{t-k_2}^2 \varepsilon_{t-l_2}^2 \varepsilon_{t-m_2}^2\right|^{1+\delta} \\
&\leq \left(E\left|\varepsilon_{t-k_1}^4 \varepsilon_{t-l_1}^4 \varepsilon_{t-m_1}^4\right|^{1+\delta} E\left|\varepsilon_{t-k_2}^4 \varepsilon_{t-l_2}^4 \varepsilon_{t-m_2}^4\right|^{1+\delta}\right)^{1/2} \\
&\leq 2\left(\sup_{0<k,l,m\leq p} E\left|\varepsilon_{t-k}^4 \varepsilon_{t-l}^4 \varepsilon_{t-m}^4\right|^{1+\delta}\right)^{1/2}
\end{aligned}$$

where the inequality is due to the Cauchy-Schwartz inequality. Boundedness is guaranteed if $E\left|\varepsilon_{t-k}^4 \varepsilon_{t-l}^4 \varepsilon_{t-m}^4\right|^{1+\delta}$ is bounded.

$$E\left|\varepsilon_{t-k}^4 \varepsilon_{t-l}^4 \varepsilon_{t-m}^4\right|^{1+\delta} \leq \left(E\left|\varepsilon_{t-k}^8\right|^{1+\delta} E\left|\varepsilon_{t-l}^8 \varepsilon_{t-m}^8\right|^{1+\delta}\right)^{1/2}$$

by another application of Cauchy-Schwartz inequality. The first term, $E\left|\varepsilon_{t-k}^8\right|^{1+\delta}$, is guaranteed to be bounded by A6 and another application of the Cauchy-Schwartz inequality will establish that $E\left|\varepsilon_{t-k}^{16}\right|^{1+\delta}$ being bounded is sufficient to establish that $E\left|\varepsilon_t^2 \widetilde{\varepsilon}_{t-1} - E\left(\varepsilon_t^2 \widetilde{\varepsilon}_{t-1} \left|\mathcal{F}_{t-1}\right.\right)\right|^{1+\delta} \leq$

$\Delta < \infty$ which enables the application of Andrew's LLN to A_1. $T^{-1}\sum_{t=1}^{T} A_2 \xrightarrow{p} 0$ by assumption A5 which is sufficient for $\overline{Z}_{T,i_6} \xrightarrow{p} 0$. Repeated application of Lemma AH then shows that $\overline{Z}_T \xrightarrow{p} 0$ which finally establishes the applicability of a LLN.

The proof for $\widetilde{\varepsilon}_{t-1}$ being a cross product of order 4 or 5 is completely analogous to that shown here and it is easy to see that the moment restrictions in A10 are more than sufficient. ■

Lemma 11 *Given Assumptions A,* $\left\{\mathbf{y}_{t-1}'\varepsilon_t\varepsilon_t'\boldsymbol{\lambda}_t\right\}$ *satisfies the UWLLN and UC assumption.*

Proof. Given the definition of $\boldsymbol{\lambda}_t$ it is apparent that there will be terms of the form $\{\varepsilon_t^2\widetilde{\varepsilon}_{t-1}\}$ where $\widetilde{\varepsilon}_{t-1}$ is a cross product of either order 3 or 4. Using the same arguments as in Lemma A5a2 and assumption A4 and A5 it can be shown that $\left\{\mathbf{y}_{t-1}'\varepsilon_t\varepsilon_t'\boldsymbol{\lambda}_t\right\}$ can be applied to a WLLN. ■

Lemma 12 *Under Assumptions A* $\left(\boldsymbol{\lambda}_t^0 - \boldsymbol{\mu}_t^0\mathbf{B}_T^0\right)'\varepsilon_t^0$ *follows a CLT.*

Proof. The sequence $\left\{\boldsymbol{\lambda}_t^0 - \boldsymbol{\mu}_t^0\mathbf{B}_T^0\right\} = \{\boldsymbol{\lambda}_t - \mathbf{y}_{t-1}\mathbf{B}_T\}$ is a series of regression residuals calculated on the basis of information contained in y_{t-i}, where $i \geq 1$ and therefore, following Lemma DGP, information contained in ε_{t-i}, with i as before. By assumption A1, this implies that $\left(\boldsymbol{\lambda}_t^0 - \boldsymbol{\mu}_t^0\mathbf{B}_T^0\right)'\varepsilon_t^0$ is again a m.d.s. and he applicability of a CLT can be established as in Gonçalves and Kilian (Lemma A1). This, however, requires the stronger moment condition A6, due to the presence of third-order cross products in $\boldsymbol{\lambda}_t$. ■

This proves the main Lemma of the paper.

Proof. Lemma 2. Wooldridge (1990, pp 41) spells out the following assumptions required for Theorem 1 to be applicable.

(i) $\Phi \subset R^p$ is compact and has nonempty interior.

(ii) $\boldsymbol{\beta}_0 \in int\,(\Phi)$.

These are very standard regularity conditions which are routinely assumed for the linear GARCH processes under consideration.

(iii) (a) $\{\varepsilon_t\,(y_t, \mathbf{y}_{t-1}, \boldsymbol{\beta}) : \boldsymbol{\beta} \in \Phi\}$ is a sequence of scalar functions such that $\varepsilon_t\,(\cdot, \boldsymbol{\beta})$ is Borel measurable for each $\boldsymbol{\beta} \in \Phi$ and $\varepsilon_t\,(y_t, \mathbf{y}_{t-1}, \cdot)$ is continuously differentiable on the interior of Φ

for all $y_t, \mathbf{y}_{t-1}$, $t = 1, 2, \ldots$. Assumption A assumes measurability of ε_t wrt to the σ-field $\mathcal{F}_{t-1}$ and continuous differentiability is apparent from the linear specification of $\varepsilon_t(\cdot) = y_t - m_t(\cdot)$.

(iii) (b) Define $\boldsymbol{\mu}_t(\mathbf{y}_{t-1}, \boldsymbol{\beta}) \equiv E\left[\partial/\partial\boldsymbol{\beta}\left(\varepsilon_t\left(y_t, \mathbf{y}_{t-1}, \boldsymbol{\beta}_0\right)\right)|\mathbf{y}_{t-1}\right]$ for all $\boldsymbol{\beta}_0 \in int(\Phi)$. Assume that $\boldsymbol{\mu}_t(\mathbf{y}_{t-1}, \cdot)$ is continuously differentiable on the interior of Φ for all $\mathbf{y}_{t-1}$, $t = 1, 2, \ldots$. Due to the linearity of $m_t(\cdot)$, it follows that $\boldsymbol{\mu}_t\left(\mathbf{y}_{t-1}, \widehat{\boldsymbol{\beta}}\right) = \mathbf{y}_{t-1}$ and hence this condition is trivially fulfilled.

(iii) (c) In these assumptions Wooldridge imposes requirements on a weighting vector, C_t, which is set to be identical to 1 for all observations in the current application. All imposed assumptions regards measurability, symmetry, positive semidefiniteness and differentiability are hence fulfilled.

(iii) (d) $\{\boldsymbol{\lambda}_t(\mathbf{y}_{t-1}, \boldsymbol{\beta}) : \boldsymbol{\beta} \in \Phi\}$ is a sequence of $1 \times q$ vectors satisfying the measurability requirements and $\boldsymbol{\lambda}_t(\mathbf{y}_{t-1}, \cdot)$ is differentiable on $int(\Phi)$ for all $\mathbf{y}_{t-1}$, $t = 1, 2, \ldots$. Given that for the V23 test $\boldsymbol{\lambda}_t\left(\mathbf{y}_{t-1}, \widehat{\boldsymbol{\beta}}\right)$ is the vector of unique second and third order cross-products of elements in $\mathbf{y}_{t-1}$, and therefore is independent of any parameter values, the latter condition is again trivially given. For the former it is important to note that as per process assumptions y_t is measurable and it follows from Theorem 3.33 in Davidson (1994) that it's cross products are measurable as well.

(iv) (a) $T^{1/2}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) = O_p(1)$. This assumption is routinely fullfilled by an ML estimate such as the usual OLS estimate $\widehat{\boldsymbol{\beta}}$.

(iv) (b) is not required due to the absence of any nuisance parameters.

(v) (a) $\left\{\boldsymbol{\mu}_t(\boldsymbol{\beta})' \boldsymbol{\mu}_t(\boldsymbol{\beta})\right\} = \left\{\mathbf{y}_{t-1}'\mathbf{y}_{t-1}\right\}$ and $\left\{\boldsymbol{\mu}_t(\boldsymbol{\beta})' \boldsymbol{\lambda}_t(\boldsymbol{\beta})\right\} = \left\{\mathbf{y}_{t-1}'\boldsymbol{\lambda}_t\right\}$ satisfy the UWLLN and UC conditions. See Lemma 8 and 9.

(v) (b) $T^{-1}\sum_{t=1}^{T} E\left[\boldsymbol{\mu}_t^0(\boldsymbol{\beta})' \boldsymbol{\mu}_t^0(\boldsymbol{\beta})\right] = T^{-1}\sum_{t=1}^{T} \mathbf{y}_{t-1}'\mathbf{y}_{t-1}$ is uniformly positive definite. This is an empirical variance covariance matrix and therefore is positive definite.

(vi) (a) $\left\{\boldsymbol{\mu}_t(\boldsymbol{\beta})' \partial/\partial\boldsymbol{\beta}\left(\varepsilon_t(\boldsymbol{\beta})\right)\right\} = \left\{\mathbf{y}_{t-1}'\mathbf{y}_{t-1}\right\}$ (see also (v) (a)), satisfy the UWLLN and UC conditions. The remaining sequences in Wooldridge are $\{0\}$ as they involve derivatives of the form $\partial/\partial\boldsymbol{\gamma}\left(\mathbf{a}_t(\boldsymbol{\gamma})\right)$, where $\mathbf{a}_t(\cdot, \boldsymbol{\gamma}) = \mathbf{a}_t(\cdot)$ for all $t = 1, 2, \ldots$.

(vi) (b) $T^{-1/2}\sum_{t=1}^{T} E\left[\boldsymbol{\mu}_t^0(\boldsymbol{\beta})_t' \varepsilon_t(\boldsymbol{\beta}_0)\right] = T^{-1/2}\sum_{t=1}^{T} E\left[\mathbf{y}_{t-1}'\varepsilon_t(\boldsymbol{\beta}_0)\right] = O_p(1)$ is given by Assumption A1.

(vii) $\{\boldsymbol{\lambda}_t' \partial/\partial \boldsymbol{\phi} \left(\varepsilon_t\left(\boldsymbol{\beta}\right)\right)\} = \{\boldsymbol{\lambda}_t'\mathbf{y}_{t-1}\}$ (see also (v) (a)) satisfy the UWLLN and UC conditions. The remaining sequences in Wooldridge are $\{0\}$ as they involve derivatives wrt to a nuisance parameter which is not present in the case of the V23 test.

(viii) (a) $\left\{\Xi_T^0 \equiv T^{-1}\sum_{t=1}^T E\left[\left(\boldsymbol{\lambda}_t^0 - \boldsymbol{\mu}_t^0\mathbf{B}_T^0\right)'\varepsilon_t^0\varepsilon_t^{0\prime}\left(\boldsymbol{\lambda}_t^0 - \boldsymbol{\mu}_t^0\mathbf{B}_T^0\right)\right]\right\}$ is uniformly positive definite.

(viii) (b) $\Xi_T^{0-1/2}T^{-1/2}\sum_{t=1}^T\left(\boldsymbol{\lambda}_t^0 - \boldsymbol{\mu}_t^0\mathbf{B}_T^0\right)'\varepsilon_t^0 \xrightarrow{d} N\left(0, I_Q\right)$. The applicability of a CLT to $\left(\boldsymbol{\lambda}_t^0 - \boldsymbol{\mu}_t^0\mathbf{B}_T^0\right)'\varepsilon_t^0$ is proven in Lemma A8b.

(viii) (c) $\{\boldsymbol{\lambda}_t'\varepsilon_t\varepsilon_t'\boldsymbol{\lambda}_t\}$, $\left\{\mathbf{y}_{t-1}'\varepsilon_t\varepsilon_t'\boldsymbol{\lambda}_t\right\}$ and $\left\{\mathbf{y}_{t-1}'\varepsilon_t\varepsilon_t'\mathbf{y}_{t-1}\right\}$ satisfy the UWLLN and UC conditions. The UC condition is trivially fulfilled as none of the terms is dependent on $\boldsymbol{\beta}$. It was shown in GK's Theorem 3.1 that $\left\{\mathbf{y}_{t-1}'\varepsilon_t\varepsilon_t'\mathbf{y}_{t-1}\right\}$ satisfies a WLLN. See Lemma A8c1.

Proof. Theorem 3. Two things need to be established.

(a) $\Xi_T^* \xrightarrow{p} \Xi_T^0$ and

(b)

$$T^{-1/2}\sum_{t=1}^T\left(\boldsymbol{\lambda}_t^* - \boldsymbol{\mu}_t^*\mathbf{B}_T^*\right)'\varepsilon_t^* \xrightarrow{d} N\left(0, \Xi_T\right)$$

just as

$$T^{-1/2}\sum_{t=1}^T\left(\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t\mathbf{B}_T\right)'\widehat{\varepsilon}_t \xrightarrow{d} N\left(0, \Xi_T\right),$$

which is demonstrated in Wooldridge's proof to Theorem 2.1. Note that under the fixed design bootstrap scheme $\boldsymbol{\lambda}_t^* = \boldsymbol{\lambda}_t$, $\boldsymbol{\mu}_t^* = \boldsymbol{\mu}_t$ and $\mathbf{B}_T^* = \mathbf{B}_T$. To show (a)

$$\begin{aligned}
\Xi_T^* &= T^{-1}\sum_{t=1}^T E\left[\left(\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t\mathbf{B}_T\right)'\varepsilon_t^*\varepsilon_t^{*\prime}\left(\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t\mathbf{B}_T\right)\right] \\
&= T^{-1}\sum_{t=1}^T E\left[\left(\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t\mathbf{B}_T\right)'\left(\widehat{\varepsilon}_t v_t\right)\left(v_t\widehat{\varepsilon}_t\right)'\left(\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t\mathbf{B}_T\right)\right] \\
&= T^{-1}\sum_{t=1}^T E\left[\left(\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t\mathbf{B}_T\right)'\widehat{\varepsilon}_t\widehat{\varepsilon}_t'\left(\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t\mathbf{B}_T\right)\right].
\end{aligned}$$

The second line is from the definition of the wild bootstrap residuals, ignoring the asymptotically negligible rescaling. The last equality is due to the unit variance characteristic of v_t. It is shown in Wooldridge's proof to Theorem 2.1, that under his Conditions A.1, the last term converges in probability to Ξ_T^0. ■

To show (b) note that

$$\begin{aligned}
& T^{-1/2}\sum_{t=1}^{T}\left(\boldsymbol{\lambda}_t^*-\boldsymbol{\mu}_t^*\mathbf{B}_T^*\right)'\varepsilon_t^* \\
= & T^{-1/2}\sum_{t=1}^{T}\left(\boldsymbol{\lambda}_t-\boldsymbol{\mu}_t\mathbf{B}_T\right)'\widehat{\varepsilon}_t v_t \\
= & T^{-1/2}\sum_{t=1}^{T}\left(\boldsymbol{\lambda}_t-\boldsymbol{\mu}_t\mathbf{B}_T\right)'\varepsilon_t v_t - T^{-1/2}\sum_{t=1}^{T}\left(\boldsymbol{\lambda}_t-\boldsymbol{\mu}_t\mathbf{B}_T\right)'\left(\varepsilon_t-\widehat{\varepsilon}_t\right)v_t \\
= & T^{-1/2}\sum_{t=1}^{T}\left(\boldsymbol{\lambda}_t-\boldsymbol{\mu}_t\mathbf{B}_T\right)'\varepsilon_t v_t - T^{-1}\sum_{t=1}^{T}\left(\boldsymbol{\lambda}_t-\boldsymbol{\mu}_t\mathbf{B}_T\right)'\mathbf{y}_{t-1}v_t T^{1/2}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \\
\equiv & A_1^*+A_2^*
\end{aligned}$$

using $\widehat{\varepsilon}_t=\varepsilon_t-\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\mathbf{y}_{t-1}'$. It needs to be established that $A_2^*\xrightarrow{P^*}0$. This can be achieved by noting that $T^{1/2}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)$ is $O_p(1)$ and

$$T^{-1}\sum_{t=1}^{T}\left(\boldsymbol{\lambda}_t-\boldsymbol{\mu}_t\mathbf{B}_T\right)'\mathbf{y}_{t-1}v_t\xrightarrow{P^*}0$$

due to the *iid* properties and the fact that $\mathbf{y}_{t-1}$ is orthogonal to $\left(\boldsymbol{\lambda}_t-\boldsymbol{\mu}_t\mathbf{B}_T\right)$ by construction of $\mathbf{B}_T$. It remains to establish that $A_1^*\xrightarrow{dP^*}N\left(0,\Xi_T\right)$.Let $Z_t^*=\eta'\left(\boldsymbol{\lambda}_t-\boldsymbol{\mu}_t\mathbf{B}_T\right)'\varepsilon_t v_t$ where $\eta\in\mathcal{R}^q$ and $\eta'\eta=1$. Since v_t is independent of $\eta'\left(\boldsymbol{\lambda}_t-\boldsymbol{\mu}_t\mathbf{B}_T\right)'\varepsilon_t$, $E^*\left(T^{-1/2}\sum_{t=1}^{T}Z_t^*\right)=0$ and

$$Var^*\left(T^{-1/2}\sum_{t=1}^{T}Z_t^*\right)=\eta' T^{-1}\sum_{t=1}^{T}\left(\boldsymbol{\lambda}_t-\boldsymbol{\mu}_t\mathbf{B}_T\right)'\left(\boldsymbol{\lambda}_t-\boldsymbol{\mu}_t\mathbf{B}_T\right)\varepsilon_t^2\eta$$

due to v_t having unit variance. An asteriks subscript in the expectations and variance operator indicates that expectations are to be taken with respect to the bootstrap distribution. An appropriate CLT has to be applied, allowing for $\{\varepsilon_t\}$ and hence $\{Z_t^*\}$ being a martingale difference sequence. Let

$$\alpha_T^{*2}=\eta'\sum_{t=1}^{T}\left(\boldsymbol{\lambda}_t-\boldsymbol{\mu}_t\mathbf{B}_T\right)'\left(\boldsymbol{\lambda}_t-\boldsymbol{\mu}_t\mathbf{B}_T\right)\varepsilon_t^2\eta;$$

and it should be noted that $T^{-1}\alpha_T^{*2}\xrightarrow{P}\Xi_T$. If for some $r>1$

$$\alpha_T^{*-2r}\sum_{t=1}^{T}E^*\left|Z_t^*\right|^{2r}\xrightarrow{P^*}0 \tag{15}$$

then $\alpha_T^{*-1} \sum_{t=1}^T Z_t^* \overset{dP^*}{\to} N(0,1)$ (this is as in GK, Proof to Theorem 3.3). If the latter result is valid then Slutsky's Theorem (Davidson, 1994, Th. 18.10) can be used to establish that $T^{-1/2} \sum_{t=1}^T Z_t^* \overset{dP^*}{\to} N(0, \eta' \Xi_T \eta)$, which is sufficient to show that $A_1^* \overset{P^*}{\to} N(0, \Xi_T)$. It remains to be checked whether the Lyapounov condition in (15) is satisfied.

$$\begin{aligned}
\alpha_T^{*-2r} \sum_{t=1}^T E^* \left| Z_t^* \right|^{2r} &= \alpha_T^{*-2r} \sum_{t=1}^T E^* \left| \eta' \left(\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t \mathbf{B}_T \right)' \varepsilon_t v_t \right|^{2r} \\
&= \alpha_T^{*-2r} \sum_{t=1}^T \left| \eta' \left(\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t \mathbf{B}_T \right)' \varepsilon_t \right|^{2r} E^* \left| v_t \right|^{2r} \\
&= \left(\frac{\alpha_T^{*2}}{T} \right)^{-r} T^{-r} \sum_{t=1}^T \left| \eta' \left(\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t \mathbf{B}_T \right)' \varepsilon_t \right|^{2r} E^* \left| v_t \right|^{2r}.
\end{aligned}$$

As the first term converges to Ξ_T (see above) it is necessary to establish that

$$T^{-r} \sum_{t=1}^T \left| \eta' (\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t \mathbf{B}_T)' \varepsilon_t \right|^{2r} E^* \left| v_t \right|^{2r} \to 0.$$

$E^* \left| v_t \right|^{2r} \leq \Delta < \infty$ by assumption for the bootstrap random process. If further $|\eta'(\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t \mathbf{B}_T)'\varepsilon_t|^{2r} \leq \Delta < \infty$ is valid, the sum will be bounded and the multiplication with T^{-r} where $r > 1$ ensures convergence to 0. In order to establish $\left| \eta' \left(\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t \mathbf{B}_T \right)' \varepsilon_t \right|^{2r} \leq \Delta < \infty$ it is necessary to recall that $\boldsymbol{\mu}_t = \mathbf{y}_{t-1}$ and that $\boldsymbol{\lambda}_t$ is a vector of all unique second- and third order cross-products of elements in $\mathbf{y}_{t-1}$. From Lemma 3 it is obvious that the highest order of ε_{t-j} to appear in $(\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t \mathbf{B}_T)'$ is ε_{t-j}^3 and therefore assumption A10 ensures $\left| \eta' \left(\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t \mathbf{B}_T \right)' \varepsilon_t \right|^{2r} \leq \Delta < \infty$.

Proof. Lemma 4. This proof is a straight application of results established in He and Teräsvirta (1999) for GARCH processes of the type $\varepsilon_t = z_t h_t$ where $h_t^k = g(z_{t-1}) + c(z_{t-1}) h_{t-1}^k$. This general specification simplifies to the GARCH(1,1) model for $k = 2$, $g(z_{t-1}) = \alpha_0$ and $c(z_{t-1}) = \beta + \alpha_1 z_{t-1}^2$. Their Theorem 1 establishes that the $km-$th unconditional moment of ε_t exists iff $E\left(c(z_{t-1})^m\right) < 1$. For the GARCH(1,1) model we have $k = 2$, and therefore the existence of the 4th [8th] {16th} moment requires $E\left(c(z_{t-1})^2\right) < 1$ [$E\left(c(z_{t-1})^4\right) < 1$] {$E\left(c(z_{t-1})^8\right) < 1$}.

The Lemma follows immediately from:

$$E\left(c(z_{t-1})^2\right) = \left(\beta + \alpha_1 z_{t-1}^2\right)^2 = \beta^2 + 2\beta\alpha_1 E\left(z_{t-1}^2\right) + \alpha_1^2 E\left(z_{t-1}^4\right)$$

$$\begin{aligned} E\left(c\left(z_{t-1}\right)^{4}\right) &= \left(\beta^{2}+2\beta\alpha_{1}E\left(z_{t-1}^{2}\right)+\alpha_{1}^{2}E\left(z_{t-1}^{4}\right)\right)^{2} \\ &= \left(\beta^{2}+2\beta\alpha_{1}E\left(z_{t-1}^{2}\right)+\alpha_{1}^{2}E\left(z_{t-1}^{4}\right)\right)\left(\beta^{2}+2\beta\alpha_{1}E\left(z_{t-1}^{2}\right)+\alpha_{1}^{2}E\left(z_{t-1}^{4}\right)\right) \\ &= \beta^{4}+4\beta^{3}\alpha_{1}E\left(z_{t-1}^{2}\right)+6\beta^{2}\alpha_{1}^{2}E\left(z_{t-1}^{4}\right)+4\beta\alpha_{1}^{3}E\left(z_{t-1}^{6}\right)+\alpha_{1}^{4}E\left(z_{t-1}^{8}\right). \end{aligned}$$

$$\begin{aligned} E\left(c\left(z_{t-1}\right)^{8}\right) &= \left(\beta^{4}+4\beta^{3}\alpha_{1}E\left(z_{t-1}^{2}\right)+6\beta^{2}\alpha_{1}^{2}E\left(z_{t-1}^{4}\right)+4\beta\alpha_{1}^{3}E\left(z_{t-1}^{6}\right)+\alpha_{1}^{4}E\left(z_{t-1}^{8}\right)\right)^{2} \\ &= \beta^{8}+4\beta^{7}\alpha_{1}E\left(z_{t-1}^{2}\right)+6\beta^{6}\alpha_{1}^{2}E\left(z_{t-1}^{4}\right)+4\beta^{5}\alpha_{1}^{3}E\left(z_{t-1}^{6}\right)+\beta^{4}\alpha_{1}^{4}E\left(z_{t-1}^{8}\right) \\ &\quad +4\beta^{7}\alpha_{1}E\left(z_{t-1}^{2}\right)+16\beta^{6}\alpha_{1}^{2}E\left(z_{t-1}^{4}\right)+24\beta^{5}\alpha_{1}^{3}E\left(z_{t-1}^{6}\right)+16\beta^{4}\alpha_{1}^{4}E\left(z_{t-1}^{8}\right) \\ &\quad +4\beta^{3}\alpha_{1}^{5}E\left(z_{t-1}^{10}\right)+6\beta^{6}\alpha_{1}^{2}E\left(z_{t-1}^{4}\right)+24\beta^{5}\alpha_{1}^{3}E\left(z_{t-1}^{6}\right)+36\beta^{4}\alpha_{1}^{4}E\left(z_{t-1}^{8}\right) \\ &\quad +24\beta^{3}\alpha_{1}^{5}E\left(z_{t-1}^{10}\right)+6\beta^{2}\alpha_{1}^{6}E\left(z_{t-1}^{12}\right)+4\beta^{5}\alpha_{1}^{3}E\left(z_{t-1}^{6}\right)+16\beta^{4}\alpha_{1}^{4}E\left(z_{t-1}^{8}\right) \\ &\quad +24\beta^{3}\alpha_{1}^{5}E\left(z_{t-1}^{10}\right)+16\beta^{2}\alpha_{1}^{6}E\left(z_{t-1}^{12}\right)+4\beta\alpha_{1}^{7}E\left(z_{t-1}^{14}\right)+\beta^{4}\alpha_{1}^{4}E\left(z_{t-1}^{8}\right) \\ &\quad +4\beta^{3}\alpha_{1}^{5}E\left(z_{t-1}^{10}\right)+6\beta^{2}\alpha_{1}^{6}E\left(z_{t-1}^{12}\right)+4\beta\alpha_{1}^{7}E\left(z_{t-1}^{14}\right)+\alpha_{1}^{8}E\left(z_{t-1}^{16}\right) \\ &= \beta^{8}+8\beta^{7}\alpha_{1}E\left(z_{t-1}^{2}\right)+28\beta^{6}\alpha_{1}^{2}E\left(z_{t-1}^{4}\right)+56\beta^{5}\alpha_{1}^{3}E\left(z_{t-1}^{6}\right)+70\beta^{4}\alpha_{1}^{4}E\left(z_{t-1}^{8}\right) \\ &\quad +56\beta^{3}\alpha_{1}^{5}E\left(z_{t-1}^{10}\right)+28\beta^{2}\alpha_{1}^{6}E\left(z_{t-1}^{12}\right)+8\beta\alpha_{1}^{7}E\left(z_{t-1}^{14}\right)+\alpha_{1}^{8}E\left(z_{t-1}^{16}\right) \end{aligned}$$

Moment existence conditions for ARCH(1) models can be regarded as special cases of these conditions with $\beta = 0$. ■ ■

Proof. Lemma 5. Substituting the parameter values for ARCH1, ARCH2, GARCH1, GARCH2 and GARCH3 with standard normal innovations, z_t, for which $E\left(z_{t-1}^{2}\right) = 1$, $E\left(z_{t-1}^{4}\right) = 3$ and $E\left(z_{t-1}^{6}\right) = 15$, $E\left(z_{t-1}^{8}\right) = 105$, $E\left(z_{t-1}^{10}\right) = 945$, $E\left(z_{t-1}^{12}\right) = 10,395$, $E\left(z_{t-1}^{14}\right) = 135,135$ and $E\left(z_{t-1}^{16}\right) = 2,027,025$ into the moment existence condition established in Lemma 4, immediately shows the result in the Lemma.

The results for the processes with innovations $z_{t-1} \sim t\left(5\right)$ follows from the nonexistence of even moments with order larger than the degrees of freedom.

The results for GARCH processes driven by $\chi^{2}\left(2\right)$ innovations depends on the moments of the innovation distribution. As higher moments of the gamma distribution are easily derived it is useful to make use of the following relation between chi-square and gamma distributed random variables: $1/2 \cdot \chi^{2}\left(n\right) = \gamma\left(n/2\right)$ and hence $1/2 \cdot \chi^{2}\left(2\right) = \gamma\left(1\right)$. Let $\widetilde{\omega}_t = 1/2\ \omega_t$, where $\omega_t \sim \chi^{2}\left(2\right)$. As $\widetilde{\omega}_t \sim \gamma\left(1\right)$, $E\left(\widetilde{\omega}_t^{r}\right) = \Gamma\left(r+1\right)/\Gamma\left(1\right)$ and with $\Gamma\left(1\right) = 1$ it follows that

$E\left(\omega_t^r\right) = 2^r\ \Gamma\left(r+1\right)$. Recognising that for integer n the gamma function $\Gamma\left(n\right) = \left(n-1\right)!$, it is easy to derive the non-central moments for ω_t from $E\left(\omega_t^r\right) = 2^r \cdot (r)!$. The innovations used in this paper are, however, not random variables ω_t but rather $z_t = \omega_t - E\left(\omega_t\right)$ and it is therefore required to derive the central moments rather than the non-central moments. The relation between central and non-central moments is (Abramowitz and Stegun, 1972, p.928):

$$E_c\left(\omega_t^r\right) = \sum_{j=0}^{r} \left(\begin{array}{c} r \\ j \end{array}\right) (-1)^{r-j}\, E\left(\omega_t^j\right) E\left(\omega_t\right)^{r-j}.$$

This yields the following higher moments for $z_t : E\left(z_{t-1}^2\right) = 4$, $E\left(z_{t-1}^4\right) = 144$, $E\left(z_{t-1}^6\right) = 16,960$ and $E\left(z_{t-1}^8\right) = 3,797,248$. The random deviates used here are standardised, such that $E\left(z_{t-1}^2\right) = 1$. Applying these results to the moment existence condition in Lemma 4 establishes that the 8th and higher moment do not exist and the result of this Lemma follows. ■

References

Abramowitz, M. and Stegun, I. A. (1972) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th printing. New York: Dover.

Barnett W.A., Gallant A.R., Hinich M.J., Jungeilges J.A., Kaplan D.T. and Jensen M.J. (1997) "A single-blind controlled competition among tests for nonlinearity and chaos" *Journal of Econometrics*, **82**, 157-192.

Bergström P. (1999) "Bootstrap methods and applications in econometrics - a brief survey" mimeo, Department of Economics, Uppsala University.

Bickel P.J. and Freedman D.A. (1981) "Some asymptotic theory for the bootstrap" *Annals of Statistics*, **9**, 1196-1217.

Bollerslev T. (1986) "Generalised autoregressive conditional heteroskedasticity", *Journal of Econometrics*, **31**, 307-327.

Bose A. (1988) “Edgeworth correction by bootstrap an autoregressions” *The Annals of Statistics*, **16**, 1709-1722.

Brock W. A., Dechert W. D., Scheinkman J. A. and LeBaron, B. (1996) “A Test for Independence Based on the Correlation Dimension” *Econometric Reviews*, **15**, 3, 197-235.

Brockwell P.J. and Davis R.A. (1991) *Time Series: Theory and Methods*, Springer-Verlag: New York.

Dahl C.M. (1999) “An investigation of tests for linearity and the accuracy of flexible non-linear inference”, Working Paper 1999-8, Department of Economics, University of Aarhus.

Davidson J. (1994) *Stochastic Limit Theory*, Oxford University Press, Oxford.

Davidson J. (2000) *Econometric Theory*, Blackwell Publishers, Oxford.

Davidson R. and Flachaire E. (2000) “The wild bootstrap. Tamed at last”, mimeo, GREQAM, Université de la Méditernée.

Davidson R. and MacKinnon J.G. (1985) “Heteroskedasticity-robust tests in regression directions” *Annals de l”INSEE*, **59/60**, 183-218.

Davidson R. and MacKinnon J.G. (1993) *Estimation and Inference in Econometrics*, Oxford University Press, Oxford.

Deo R. (2000) “Spectral tests of the martingale hypothesis under conditional heteroscedasticity”, *Journal of Econometrics*, **99**, 291-315.

Elliot, G., Komunjer, I. and Timmerman, A. (2003) “Estimating loss function parameters”, Unpublished Discussion Paper, University of California San Diego.

Engle R. (1982) “Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation”, *Econometrica*, **50**, 987-1007.

Flachaire E. (1999) “A better way to bootstrap pairs”, *Economics Letters*, **64**, 257-262.

Godfrey L.G. and Orme C.D. (2001) “Significance levels of heteroskedasticity-robust tests for specification and misspecification: some results on the use of wild bootstraps” unpublished manuscript.

Godfrey L.G. and Orme C.D. (2004) “Controlling the finite sample significance levels of heteroskedasticity-robust tests of several linear restrictions on regression coefficients”, *Economics Letters*, **82**, 281-287.

Gonçalves S. and Kilian L. (2004) “Bootstrapping autoregressions with conditional heteroskedasticity of unknown form”, *Journal of Econometrics*, **123**, 89-120.

Gonçalves S. and White H. (2004) “Maximum likelihood and the bootstrap for nonlinear dynamic models”, *Journal of Econometrics*, **119**, 199-220.

Gonçalves S. and White H. (2001) “The bootstrap of the mean for dependent heterogeneous arrays”, Econometric Theory, **18**, 1367-1384.

Granger C.W.J. and Teräsvirta T. (1993) *Modelling Nonlinear Economic Relationships*, Oxford University Press: Oxford.

Halunga A. (2005) “Wild Bootstrap Tests for Autocorrelation in the Presence of Possible Conditional Heteroskedasticity of Unknown Form”. Mimeo. University of Manchester.

He C. and Teräsvirta T. (1999) “Properties of moments of a family of GARCH processes”, *Journal of Econometrics*, **92**, 173-192.

Hinich M.J. (1982) “Testing for gaussianity and linearity of stationary time series”, *Journal of Time Series Analysis*, **3**, 169-176.

Horowitz J.L. (1997) “The bootstrap” mimeo, Department of Economics, University of Iowa, to appear in Handbook of Econometrics, Volume V.

Kim, D.H., Osborn, D.R. and Sensier, M. (2002) “Nonlinearity in the Fed’s monetary policy rule”, Unpublished Discussion Paper, University of Manchester.

Künsch H.R. (1989) "The jacknife and the bootstrap for general stationary observations" *The Annals of Statistics*, **17**, 1217-1241.

Lee, T., White, H. and Granger, C.W.J., (1993) "Testing for neglected nonlinearity in time series models", *Journal of Econometrics*, **56**, 269-290.

Li H. and Maddala G.S (1996) "Bootstrapping time series models" *Econometric Reviews*, **15**, 115-195 (incl. discussion).

Liu R.Y. (1988) "Bootstrap procedures under some non-iid models", *The Annals of Statistics*, **16**, 1696-1708.

MacKinnon J.G. (1992) "Model specification tests and artificial regressions" *Journal of Economic Literature*, **30**, 102-146.

Mammen E. (1993) "Bootstrap and wild bootstrap for high dimensional linear models", *The Annals of Statistics*, **21**, 255-285.

McLoed, A. I. and Li, W. K. (1983) "Diagnostic checking ARMA time series models using squared-residual autocorrelations" *Journal of Time Series Analysis*, **4**, 4, 269-273.

Patterson D.M. and Ashley R.A. (2000) *A Nonlinear Time Series Workshop: A Toolkit for Detecting and Identifying Nonlinear Seriel Dependence*, Kluwer Academic Publishers: Boston.

Poltis D.N. and Romano J.P. (1994) "The stationary bootstrap" *Journal of the American Statistical Association*, **89**, 1303-1313.

Ramsey, J.B. (1969) "Tests for specification errors in classical linear least squares regression analysis" *Journal of the Royal Statistical Society B*, **31**, 350-371.

Sin C.-Y. and White H. (1996) "Information criteria for selecting possibly misspecified parametric models" *Journal of Econometrics*, **71**, 207-225.

Teräsvirta T., Lin C.F. and Granger C.W.J. (1993) "Power of the neural network linearity test", *Journal of Time Series Analysis*, **14**, 209-220.

Teräsvirta T., Tjøstheim D. and Granger C.W.J. (1994) "Aspects of modelling nonlinear time series", in: Engle R.F. and McFadden D.L., Handbook of Econometrics, Vol.IV, Elsevier Science.

Tsay, R.S. (2002) *Analysis of Financial Time Series*, John Wiley & Sons: New York.

White H. (1984) *Asymptotic Theory for Econometricians*, Academic Press: Orlando.

White H. (1989) "An additional hidden unit test for neglected nonlinearity in multilayer feedworward networks" in *Proceedings of the International Joint Conference of Neural Networks, Washington, DC*, IEEE Press: New York.

Wooldridge J.M. (1990) "A unified approach to robust, regression-based specification testing" *Econometric Theory*, **6**, 17-43.

Wooldridge J.M. (1991) "On the application of robust, regression-based diagnostics to models of conditional means and conditional variances" *Journal of Econometrics*, **47**, 5-46.